CENTER FOR HEALTH ECONOMICS RESEARCH

411 Waverly Oaks Road, Suite 330
Waltham, MA 02154-8705
(781) 788-8100
(781) 788-8101 - FAX

Evaluating Alternative Risk Adjusters for Medicare

Final Report

Prepared by:

Gregory C. Pope, M.S.
Killard W. Adamache, Ph.D.
Edith G. Walsh, Ph.D.
Rezaul K. Khandker, Ph.D.

Center for Health Economics Research

Prepared for:

Health Care Financing Administration

March 26, 1998

Gregory C. Pope

Gregory C. Pope, M.S.
Project Director

Debra A. Dayhoff

Debra A. Dayhoff, Ph.D.
Scientific Reviewer

This work was performed under HCFA Cooperative Agreement No. 17-C-90316/1-02, Sherry Terrell, Project Officer. The authors thank Sherry Terrell, Cynthia Tudor, Mel Ingber and other HCFA staff for helpful comments. Randall Ellis also provided useful comments. However, opinions and any errors are the responsibility of the authors alone. No endorsement by HCFA or DHHS is intended or should be inferred.

Table of Contents

Table of Contents
(continued)

Table of Tables, Figures and Exhibits

Page

Executive Summary

Chapter 2

Chapter 3

Chapter 4

Table of Tables, Figures and Exhibits
(continued)

Page

Chapter 5

Table of Tables, Figures and Exhibits
(continued)

Page

Chapter 6

Abstract

This study uses multiple years of the Medicare Current Beneficiary Survey to evaluate alternative demographic, survey, and claims-based risk adjusters for Medicare capitation payment. The survey health status models have three to four times the predictive power of the demographic models. The risk adjustment models derived from claims diagnoses have greater predictive power than the survey models. No single model predicts average expenditures well for all beneficiary subgroups of interest, suggesting a combined model may be appropriate. More data are needed to obtain stable estimates of model parameters. Advantages and disadvantages of alternative risk adjusters are discussed.

Executive Summary

1.0 Introduction

Although it is a goal of the Medicare program to enroll more of its beneficiaries in managed care programs, as of January 1997, only about 13 percent were enrolled in Health Maintenance Organizations (HMOs) or Competitive Medical Plans (1997 HCFA Statistics). Medicare's managed-care enrollment shortfall has been attributed in part to the inadequacy of its current payment formula for HMOs--the Adjusted Average Per Capita Cost (AAPCC)--in accounting for expenditure differences among beneficiaries. The AAPCC considers only sociodemographic factors (age, sex, employment status, welfare status, and institutional status), location (county of residence), and reason for Medicare eligibility (aged, disabled, or ESRD). Many studies have shown that the AAPCC factors inadequately predict medical expenditures, creating inequities among HMOs that enroll healthier or sicker beneficiaries, and also large financial incentives for HMOs to try to attract healthier beneficiaries. Numerous proposals have been advanced to incorporate additional factors into the AAPCC, such as health status, prior medical care use, diagnoses, and medical risk factors. The additional factors that could be added to the AAPCC generally are available from two sources: surveys of beneficiaries or medical claims data. The 1997 Balanced Budget Act requires Medicare to adopt a health-status

risk-adjusted capitated payment methodology by January 1, 2000. The new payment methodology is likely to incorporate health status factors collected from either surveys or claims, or both. It is thus of interest to examine the merits of alternative survey and claims-based risk adjusters for the Medicare population.

The largest survey of the Medicare population currently available is the Medicare Current Beneficiary Survey (MCBS). This study employs three years of the MCBS and associated claims data to evaluate alternative survey and claims-based risk adjusters on a common sample. With the requirement that Medicare claims for ambulatory patients contain diagnostic codes, there have recently been substantial innovations in claims-based risk adjusters (Ellis *et al*, 1996; Weiner *et al*, 1996; Kronick *et al*, 1996). Two of the latest generation of claims-based adjusters are included in our evaluation.

2.0 Data

The MCBS is an ongoing, multi-purpose survey of a nationally representative sample of the Medicare population, including both aged and disabled enrollees who live in the community or are institutionalized. A key advantage of the MCBS for use in this study is that it links survey responses to Medicare administrative claims, enabling us to compare the performance of survey- and claims-based risk adjusters in predicting actual Medicare payments. Also, survey responses allow performance of alternative risk adjusters to be compared for groups--such as supplemental insurance status--not identifiable from Medicare administrative records.

The MCBS is a population-based survey that employs a panel design. Each round of the MCBS includes survey data and Medicare claims data collected for the same individuals. The claims data include diagnostic codes and the Medicare expenditures associated with each claim. For this study, we used data from Rounds 1, 4 and 7, in the years 1991, 1992, and 1993, respectively. We used Round 1 (1991) survey data and claims-based diagnostic groups as predictors of Round 4 (1992) expenditures, and Round 4 (1992) survey and claims-based diagnostic groups as predictors of Round 7 (1993) expenditures.

3.0 Study Design

We exploit the longitudinal nature of the MCBS by estimating our models using 1991 survey and claims data to predict 1992 expenditures. We then validate the models using 1992 survey and claims data to predict 1993 expenditures. Two years of data are necessary for both estimation and validation because we are evaluating prospective risk adjustment models that use beneficiary characteristics to predict expenditures in the subsequent year. That is, using the regression parameters from the 1991/1992 sample, the validation model uses 1992 Medicare beneficiary characteristics to predict 1993 Medicare expenditures. Then the predicted expenditures for 1993 are compared to the actual expenditures for 1993. We emphasize performance of models on our validation rather than the estimation sample.

4.0 Risk Adjustment Models

We developed nine risk adjustment models using the information available from the MCBS. They are:

1. demographic;
2. self-rated (general) health status;
3. self-reported chronic conditions;
4. functional status;
5. short form (SF)-36 simulation;
6. comprehensive survey;
7. claims diagnoses;
8. claims diagnoses plus survey; and
9. prior use.

Each model was estimated using 1991 (Round 1) survey characteristics or claims data to predict 1992 Medicare program expenditures.

In addition to a basic age/sex model, a second demographic model incorporating additional factors, Medicaid enrollment status and institutionalization, used in Medicare's current Adjusted Average Per Capita Cost (AAPCC) methodology was estimated. A model is defined for each of three major domains of survey health status measures. These are self-rated health (also called "general" or "perceived" health status), self-reported chronic conditions, and functional status. Our functional status variable is a count of the number of activities of daily living (ADLs) that a respondent reports difficulty or inability performing, with an additional category for difficulty/inability with at least one instrumental activity of daily living (IADLs), but not with any ADL. These measures, along with limitations walking 2-3 blocks or lifting 10 pounds, are also combined into a comprehensive survey model to analyze their joint properties.

We also developed a model simulating four of the eight scales from the SF-36 to provide a comparison to our survey models and to other work done using the SF-36 for risk adjustment. While the MCBS and SF-36 questions differ in details of wording, we were able to construct simulated scores for the Physical Functioning, General Health, Social Functioning and Role-Physical scales. These are four of the five scales that Hornbrook and Goodman (1995) found to be predictors of medical costs. Our SF-36-like scales have not been tested for equivalence to the actual SF-36 scales.

The MCBS asks respondents if a doctor has ever told them that they have any of a list of specific medical conditions (heart attack, diabetes, cancer, etc.). We measured each of these with a dichotomous yes/no variable. Conditions that were not positive and statistically significant in a preliminary model were eliminated from the final model. This same list of selected conditions was used for the comprehensive survey model combining self-rated health, functional status, and self-reported chronic conditions.

For comparison to the survey models, we included two claims-based diagnostic models using the diagnoses recorded on MCBS-linked hospital and physician claims to predict future expenditures. These are the Principal Inpatient, or PIP, and the Hierarchical Coexisting Conditions, or HCC, variants of the Diagnostic Cost Group or DCG models, which are described in Ellis *et al* (1996). The PIPDCG model relies on principal hospital diagnoses only, whereas the DCG/HCC model uses all inpatient and ambulatory diagnoses. In addition to the PIPDCG and DCG/HCC models, we added survey measures to the claims-based scores to form two additional models. These "combined" models.

allow us to evaluate the incremental contribution of survey variables to the claims-based models. Finally, we developed a model based on prior use of medical services for comparison to the survey and claims-diagnosis models. In the prior use model, total Medicare payments for an individual in the previous year are used to predict current year payments.

5.0 Validation Results

Two measures of predictive accuracy are computed for each model, one for individuals, and one for groups. The individual measure is the R-square statistic, defined as the proportion of variation in actual payments accounted for by predicted payments. The group measure is the predictive ratio, defined as the ratio of the aggregate predicted payments for a group of beneficiaries divided by the aggregate actual payments for this group.

5.1 Predictive Accuracy, Individuals

Table ES-1 shows R-squares for the overall validation sample and by aged versus disabled. As expected, the demographic models--age/gender and AAPCC--are the least predictive models, explaining less than one percent of the variance in actual payments. Even the least powerful model incorporating health status, the functional status model, triples the predictive power of the demographic models. The comprehensive survey and SF-36-like models, which measure multiple dimensions of health status, are the most

Table ES-1

Percentage of Variation In Medicare Program Payments Explained By Alternative Risk Adjustment Models, Validation Sample

Model	Combined	Aged	Disabled
Age & Gender	0.7 %	0.7 %	0.0 %
AAPCC Like (Demographic)	0.9	0.9	0.6
Functional Status	2.5	2.7	0.1
Self-Reported Chronic Conditions	2.7	2.8	2.1
Self-Rated Health Status	3.1	3.3	0.1
Comprehensive Survey	4.2	4.3	2.6
SF-36-like[a]	4.1	4.2	1.6
Prior Use (Expenditures)	4.1	3.0	18.5
PIPDCG	5.2	4.8	9.8
PIPDCG and Survey	6.6	6.3	9.4
DCG/HCC	7.3	6.7	14.4
DCG/HCC and Survey	7.9	7.4	13.8

[a] Simulated SF-36 scales that have not been tested for equivalence with the actual scales.

NOTE: All models include age and gender. Model parameters were estimated on combined 1991/1992 aged/disabled sample.

SOURCE: 1992, 1993 Medicare Current Beneficiary Survey.

predictive survey models. But the relatively modest gain in R-square over the single-dimension survey models indicates considerable redundancy among the survey measures. The prior use model is more powerful than the individual survey measures, but less powerful than the comprehensive survey model and certainly the claims-based model.

The claims-diagnosis-based DCG models are more predictive than any of the survey models. The R-square of the DCG/HCC model exceeds that of the comprehensive survey model by about 75 percent. Adding survey functional and self-rated health status to the PIPDCG model's principal hospital diagnoses improves forecasting accuracy significantly, from 5.2 percent to 6.6 percent, or by about 25 percent. Conversely, adding the two survey measure to the DCG/HCC model's inpatient and ambulatory diagnoses results in a gain in predictive power of only 0.6 percentage points, or about eight percent. These survey variables appear to contain only a limited amount of information relevant to predicting expenditure differences among individuals not already incorporated in inpatient and ambulatory diagnoses on claims. But they do add substantial information to principal hospital diagnoses alone. Moreover, the incremental explanatory power of the survey variables may be important in "getting payment right" on average for certain policy-relevant subgroups, a subject we discuss below.

The advantage of the claims-based models differs greatly by aged versus disabled subsamples. For the disabled, the DCG/HCC model is clearly more predictive with an R-square of 14.4 percent versus only 2.6 percent for the comprehensive survey model. Among the elderly, the DCG/HCC model is still better by more than 50 percent, but the

gap in R-square is narrowed to 6.7 percent versus 4.3 percent. In addition, survey variables add more predictive power at the margin to claims diagnoses among the elderly. Prior use also does dramatically better among the disabled than the elderly. Expenditures among the disabled are more predictable, and are relatively strongly related to past expenditures and to diagnoses recorded on medical claims, making the disabled particularly suitable for claims-based risk adjustment.

5.2 Predictive Accuracy, Groups

Table ES-2 reports predictive ratios for the overall validation sample and validation subgroups. A predictive ratio closer to 1.00 indicates better prediction. A predictive ratio greater than one indicates overprediction, whereas a predictive ratio less than one indicates underprediction. For example, a predictive ratio of 1.10 indicates 10 percent overprediction whereas a ratio of 0.90 indicates 10 percent underprediction. The predictive ratios are subject to random variation because of the limited MCBS sample size. Accordingly, statistical significance of the predictive ratios (difference from 1.00) is indicated in Table ES-2.

Predictive ratios are italicized for models that include variables defining the validation group. For example, since functional status is included in the functional status, SF-36 like, comprehensive survey, and combined survey and claims models, the predictive ratios for the functional status validation groups are italicized for these models. While one would expect predictive ratios closer to one for validation groups that are defined by

Table ES-2

Predictive Ratios for Alternative Risk Adjustment Models, by Validation Subgroup
(A predictive ratio closer to 1.00 indicates better prediction)

Validation Group	Age-Gender	AAPCC[1] Like	Functional Status	Self-Reported Chronic Conditions	Self-Rated Health Status	SF-36 Like	Prior Use	Comprehensive Survey	PIPDCG	PIPDCG & Survey	DCG/HCC	DCG/HCC & Survey
Overall Sample (normalized)[2]	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Age												
0-64	1.04	1.04	0.97	1.01	1.01	0.96	1.08	0.97	1.03	1.28 ***	1.05	1.19 **
65-74	1.07 *	1.07 *	1.05	1.06	1.05	0.95	1.05	1.05	1.03	0.94	1.01	0.96
75-84	0.92	0.92 *	0.94	0.93	0.94	1.02	0.93	0.94	0.98	0.97	0.98	0.98
85+	1.00	0.99	1.03	1.01	1.01	1.05	1.01	1.03	0.96	1.06	0.99	1.05
Gender												
Female	1.04	1.04	1.05	1.05	1.04	1.04	1.05	1.05	1.01	1.03	1.01	1.02
Male	0.94	0.94	0.94	0.93 *	0.95	0.95	0.93 *	0.93	0.99	0.96	0.99	0.97
Medicare Status												
Elderly	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.97	1.00	0.98
Disabled	1.04	1.04	0.97	1.01	1.01	0.96	1.08	0.97	1.03	1.28 ***	1.05	1.19 **
Institutional Status												
Non-institutionalized	1.02	1.02	0.98	1.00	1.01	0.99	1.00	0.99	1.01	0.99	0.99	0.98
Institutionalized	0.77 ***	0.82 **	1.21 ***	1.03	0.86 *	1.15 *	0.96	1.17 **	0.88	1.16 **	1.12	1.27 ***
Self Rated Health Status												
poor	0.49 ***	0.50 ***	0.69 ***	0.68 ***	0.94	0.88 *	0.72 ***	0.93	0.64 ***	0.94	0.76 ***	0.94
fair	0.77 ***	0.78 ***	0.92	0.93	1.00	1.06	0.87 **	1.02	0.85 ***	1.01	0.96	1.03
good	0.97	0.97	0.96	0.96	0.93	0.94	0.95	0.93	0.96	0.94	0.98	0.94
very good	1.42 ***	1.40 ***	1.20 ***	1.22 ***	1.05	1.02	1.23 ***	1.06	1.28 ***	1.03	1.13 **	1.02
excellent	2.15 ***	2.11 ***	1.69 ***	1.64 ***	1.31 ***	1.25 ***	1.77 ***	1.28 ***	1.88 ***	1.29 ***	1.47 ***	1.25 ***
Functional Status[3]												
5-6 ADLs	0.58 ***	0.61 ***	1.10	0.82 ***	0.76 ***	1.02	0.85 **	1.08	0.72 ***	1.06	0.88 *	1.08
3-4 ADLs	0.66 ***	0.67 ***	0.97	0.84 **	0.83 **	1.06	0.83 **	0.94	0.74 ***	0.94	0.85 *	0.95
1-2 ADLs	0.79 ***	0.80 ***	1.03	0.88 **	0.89 **	0.99	0.85 ***	1.03	0.85 ***	1.03	0.90 **	1.03
IADLs only	1.04	1.05	0.95	1.04	1.04	1.04	1.05	0.95	1.06	0.97	1.04	0.96
None	1.44 ***	1.41 ***	0.97	1.21 ***	1.23 ***	0.97	1.22 ***	0.99	1.30 ***	0.98	1.16 ***	0.98
Elderly helped with 3+ ADLs	0.55 ***	0.58 ***	0.71 ***	0.91	0.82 ***	0.94	0.84 ***	0.97	0.70 ***	0.96	0.88 *	1.00
Chronic Conditions												
Any Chronic Condition	0.95 *	0.95 *	0.97	0.99	0.97	0.98	0.98	1.00	0.97	0.99	0.99	1.00
Arteriosclerosis	0.74 ***	0.74 ***	0.87 **	1.08	0.86 **	0.93	0.85 ***	1.09	0.84 ***	0.94	0.95	1.00
Heart Attack	0.61 ***	0.61 ***	0.68 ***	0.96	0.71 ***	0.74 ***	0.77 ***	0.97	0.74 ***	0.79 ***	0.86 **	0.89 *
Angina	0.62 ***	0.63 ***	0.70 ***	0.88 *	0.73 ***	0.76 ***	0.75 ***	0.90 *	0.73 ***	0.80 ***	0.86 **	0.89 *

Table ES-2 (continued)

Predictive Ratios for Alternative Risk Adjustment Models, by Validation Subgroup
(A predictive ratio closer to 1.00 indicates better prediction)

Validation Group	Age-Gender	AAPCC[1] Like	Functional Status	Self-Reported Chronic Conditions	Self-Rated Health Status	SF-36 Like	Prior Use	Compre-hensive Survey	PIPDCG	PIPDCG & Survey	DCG/HCC	DCG/HCC & Survey
Chronic Conditions (continued)												
Other Heart Conditions	0.70 ***	0.71 ***	0.78 ***	_0.95_	0.80 ***	0.83 ***	0.83 ***	_0.96_	0.82 ***	0.88 **	0.94	0.96
Hypertension	0.89 ***	0.90 ***	0.93 **	0.97	0.94	0.96	0.93 **	0.99	0.92 **	0.95	0.96	0.97
Stroke	0.73 ***	0.75 ***	0.95	1.01	0.87 **	1.01	0.93	1.06	0.85 ***	1.00	1.01	1.09 *
High Cost Cancer	0.80 *	0.81 *	0.87	_1.18_	0.91	0.95	0.96	_1.20_ *	0.97	1.03	1.15	1.17
Low Cost Cancer	0.89 *	0.88 *	0.93	0.96	0.93	0.94	1.00	0.96	0.94	0.96	1.04	1.05
Skin Cancer	0.97	0.95	0.97	1.01	0.98	0.97	0.99	1.00	0.99	0.97	1.02	1.00
Diabetes	0.62 ***	0.64 ***	0.71 ***	_0.95_	0.73 ***	0.77 ***	0.73 ***	_0.96_	0.72 ***	0.80 ***	0.95	0.98
Rheumatoid Arthritis	0.77 ***	0.78 ***	0.91	0.92	0.90	0.98	0.84 ***	1.00	0.81 ***	0.93	0.87 **	0.94
Osteoarthritis	0.93 *	0.94 *	0.99	1.01	0.99	1.03	0.97	1.04	0.95	1.00	0.98	1.00
Osteoporosis	0.95	0.96	1.17 **	_1.39_ ***	1.08	1.22 ***	1.14 **	_1.40_ ***	1.00	1.17 ***	1.10	1.20 ***
Mental Retardation	1.33 **	1.47 ***	1.39 ***	1.19	1.10	1.21	1.31 **	1.09	1.34 **	1.55 ***	1.35 **	1.47 ***
Dementia	0.94	0.98	1.39 ***	1.21 **	1.12	1.42 ***	1.08	1.38 ***	1.01	1.33 ***	1.21 **	1.38 ***
Mental Disorders	0.81 **	0.84 *	0.89	0.88	0.88	0.93	0.91	0.91	0.86 *	1.02	0.93	1.01
Hip Fracture	0.92	0.94	1.20 **	1.07	1.00	1 20 **	1.12	1.19 **	1.01	1.21 **	1.13	1.24 ***
Parkinson's Disease	0.76	0.77	1.04	_1.32_	0.93	1.08	0.87	_1.34_	0.84	1.05	1.07	1.18
Chronic Obstructive Pulmonary Disease	0.71 ***	0.71 ***	0.80 ***	_1.03_	0.84 ***	0.87 **	0.81 ***	_1.03_	0.83 ***	0.92	0.95	0.99
Partial Paralysis	0.73 ***	0.74 ***	0.98	_1.10_	0.86 *	1.02	0.93	_1.10_	0.81 ***	1.01	0.97	1.08
Amputation of arm/leg	0.57 ***	0.59 ***	0.77	_1.26_ *	0.71 **	0.85	0.86	_1.27_ *	0.72 **	0.86	0.88	0.96
Lost urine more than once per week	0.70 ***	0.72 ***	0.96	_1.06_	0.83 ***	0.99	0.88 **	_1.07_	0.79 ***	0.98	0.91 *	1.02
Expenditures, 1992												
First Quintile (lowest)	1.99 ***	1.97 ***	1.77 ***	1.64 ***	1.79 ***	1.66 ***	_1.34_ ***	1.55 ***	1.56 ***	1.41 ***	0.97	0.92
Second Quintile	1.68 ***	1.66 ***	1.54 ***	1.51 ***	1.55 ***	1.47 ***	_1.16_ *	1.44 ***	1.30 ***	1.22 **	1.13	1.10
Third Quintile	1.39 ***	1.39 ***	1.37 ***	1.39 ***	1.36 ***	1.34 ***	_1.01_	1.37 ***	1.09	1.09	1.22 ***	1.22 ***
Fourth Quintile	0.92	0.92	0.95	0.99	0.96	0.97	_0.79_ ***	0.99	0.78 ***	0.82 ***	1.02	1.03
Fifth Quintile (highest)	0.46 ***	0.47 ***	0.55 ***	0.56 ***	0.54 ***	0.60 ***	_0.97_	0.61 ***	0.86 ***	0.90 ***	0.88 ***	0.90 ***
Top 5 percent	0.31 ***	0.31 ***	0.42 ***	0.41 ***	0.39 ***	0.47 ***	_1.25_ ***	0.47 ***	0.74 ***	0.79 ***	0.86 **	0.88 *
Hospital Admissions, 1992												
no admissions	1.27 ***	1.27 ***	1.23 ***	1.22 ***	1.23 ***	1.20 ***	0.97	1.19 ***	0.99	0.98	1.02	1.02
one admission	0.63 ***	0.63 ***	0.72 ***	0.73 ***	0.70 ***	0.77 ***	1.04	0.78 ***	1.18 ***	1.21 ***	1.04	1.06
two or more admissions	0.33 ***	0.34 ***	0.41 ***	0.44 ***	0.41 ***	0.47 ***	1.07	0.49 ***	0.82 ***	0.85 **	0.86 **	0.87 *
Supplemental Insurance[4]												
Medicaid	0.74 ***	0.91	0.90 *	0.86 ***	0.83 ***	0.93	0.88 **	0.93	0.82 ***	0.98	0.93	1.01
Medicare Only	1.15	1.10	1.17	1.15	1.22 *	1.22 *	1.06	1.20 *	1.13	1.21 *	1.01	1.07
Other Supplemental Coverage	1.05 *	1.01	1.00	1.02	1.02	0.99	1.03	0.99	1.03	0.98	1.02	0.99

Table ES-2 (continued)

Predictive Ratios for Alternative Risk Adjustment Models, by Validation Subgroup
(A predictive ratio closer to 1.00 indicates better prediction)

Validation Group	Age-Gender	AAPCC[1] Like	Functional Status	Self-Reported Chronic Conditions	Self-Rated Health Status	SF-36 Like	Prior Use	Comprehensive Survey	PIPDCG	PIPDCG & Survey	DCG/HCC	DCG/HCC & Survey
Income												
≤ $15,000	0.93 **	0.95	0.96	0.95	0.96	0.98	0.95	0.98	0.95	0.99	0.96	0.98
$15,001 - $25,000	1.09	1.05	1.05	1.06	1.07	1.04	1.04	1.04	1.05	1.00	1.04	1.01
> $25,000	1.19 ***	1.15 **	1.09	1.11 *	1.07	1.03	1.14 **	1.03	1.16 **	1.02	1.11 *	1.04
Not Reported	0.67 *	0.66 *	1.01	0.88	0.79	0.98	0.82	1.00	0.77	1.00	0.92	1.04
Education												
< 12 years	0.92 **	0.94	0.95	0.95	0.98	1.00	0.93 *	0.99	0.94	0.99	0.96	0.99
= 12 years	1.08 *	1.06	1.04	1.05	1.05	1.02	1.06	1.03	1.04	1.00	1.04	1.02
> 12 years	1.09	1.06	1.02	1.02	0.98	0.96	1.08	0.96	1.08	0.98	1.00	0.96
Not Reported	0.86	0.92	1.23 *	1.06	0.96	1.20 *	0.98	1.18	0.95	1.21 *	1.15	1.30 **
Race												
White	1.01	1.00	1.01	1.01	1.00	1.00	1.01	1.00	1.01	1.00	1.01	1.00
Black	0.93	1.00	0.97	0.93	1.01	1.03	0.94	1.00	0.94	1.02	0.97	1.01
Other	0.82	0.88	0.87	0.86	0.89	0.92	0.89	0.92	0.88	0.93	0.89	0.92
Living Status												
Living Alone	0.97	0.97	0.91 *	0.94	0.94	0.90 **	0.95	0.91 *	0.95	0.92	0.95	0.93
Living with Spouse	0.85 ***	0.88 **	0.99	0.93	0.89 **	1.00	0.95	0.99	1.09 **	1.03	0.98	1.05
Living with Others	1.11 ***	1.09 **	1.06	1.08 **	1.10 **	1.07	1.06	1.07	0.91 *	1.03	1.05	1.01

NOTES:

Italics are used to indicate that the validation group is defined by variables included in the predictive model.

1. AAPCC is Adjusted Average Per Capita Cost.
2. Predictive ratios were normalized by dividing them by the predictive ratio of the overall sample.
3. ADL is activity of daily living. IADL is instrumental activity of daily living.
4. Other supplemental coverage includes individually purchased (IP), employer sponsored (ES), both IP and ES, and public coverage other than Medicaid, as well as private plans held by a small number of working elderly.

*** Predictive ratio is significantly different from 1 at the .01 level.
** Predictive ratio is significantly different from 1 at the .05 level.
* Predictive ratio is significantly different from 1 at the .10 level.

SOURCE: 1992 (Round 4) and 1993 (Round 7) Medicare Current Beneficiary Survey.

elements of the predictive model, these predictive ratios are still of interest to determine reliability, since estimation and validation are on different years. Moreover, there is no guarantee that models comprising multiple variables will predict well for validation groups defined by a single variable.

In the remainder of this section, we highlight findings for selected validation groups of particular interest.

The Institutionalized

The demographic models underpredict spending for the institutionalized, but many of the health status models overpredict spending. This indicates that nursing home residents are absolutely more expensive, but are less expensive to Medicare than community residents with the same diagnoses or functional status. Institutionalized beneficiaries may be less expensive controlling for diagnoses or functional status because of the substitution of nursing home care for the acute care services covered by Medicare.

Self-Rated Health and Functional Status

Not surprisingly, survey models including these variables predict well across validation groups. Other demographic, survey, prior use, and claims-based models are less successful. Comparison of the claims diagnosis DCG models with the combined survey/claims models indicates that survey variables can improve predictions of the claims model across health and functional status groups.

Elderly Receiving Help with ADLs

We included an "elderly receiving help with 3 or more ADLs" validation group since policy makers and providers are interested in the ability of risk adjustment methodologies to pay accurately for the more functionally impaired elders at risk of institutionalization. Only the combined models, including both the claims-based DCG and the survey measures predict accurately for this group, along with the comprehensive survey model. The self-reported chronic conditions and the SF-36-like model predict reasonably well, even though there are no functional status measures in the first, and limited functional status information in the second. All other models substantially underpredict expenditures for this group.

Prior Utilization

The models using claims information predict payments better for persons with varying levels of prior year payments than the survey or demographic models. The DCG/HCC model underpredicts by only 14 percent among the highest 5 percent prior year spenders as compared to a 53 percent underprediction by the comprehensive survey and SF-36-like models. For the lowest quintile, the DCG/HCC model mispredicts by only 3 percent, versus 64 to 79 percent overprediction by the survey variables. The PIPDCG model does nearly as well as the DCG/HCC model for the highest quintile of prior expenditures, but not as well for the top 5 percent, or for the fourth quintile, and it substantially overpredicts for the bottom 40 percent. The combined survey and claims

models do not do much better than the DCG models alone, that is, survey measures do not add much to claims diagnoses for predicting across prior expenditure quintiles. For prior-year hospital admission categories, the prior use model does best, with the two models including the DCG/HCC score a close second, and the two PIPDCG models third.

The Chronically Ill

Across groups of people reporting chronic conditions[1], the models using diagnostic information -- self-reported chronic conditions, comprehensive survey, DCG/HCC, and combined DCG/HCC/survey -- show the fewest statistically significant under- or over-predictions. (The PIPDCG model--which uses only principal inpatient diagnoses--is an exception.) The SF-36-like model also does well, despite utilizing no diagnostic information. The most important chronic condition indicators to include in survey models appear to be heart disease, diabetes, and chronic lung disease. All models overpredict expenditures for the mentally retarded, and all except the demographic models and PIPDCG overpredict for dementia. This could be due to under-provision of care to these groups, or substitution of Medicaid nursing home for Medicare expenditures.

Dual Eligibles

Medicare/Medicaid dual eligibles (identified by "Medicaid" under "Supplemental Insurance" in Table 5-2) are a group of particular interest to state and federal policy

[1] Chronic condition groups could be defined using either survey self-reports or diagnoses recorded on claims. We use self-reports.

makers. Only the combined claims and survey models predicts this group's expenditures accurately. All other models underpredict for this group, although the underpredictions of the AAPCC-like, SF-36-like, comprehensive survey and DCG/HCC models are not statistically significant. Larger sample sizes are needed to confirm these findings.

Demographic Groups

All the models predict mean expenditures reasonably well across income,[2] education, and race groups, with the exception of the age/gender model. Predicted spending is in general higher than actual spending for beneficiaries who live with individuals other than their spouse, which could reflect substitution of nursing home care for acute medical care, or underservice to these beneficiaries. Living alone, on the other hand, has a tendency to raise actual compared to predicted expenditures.

6.0 Conclusions

No one risk adjustment model is best on all empirical criteria considered in this report. The claims-diagnosis-based DCG/HCC model has greater overall predictive power than the survey, PIPDCG, and prior use models, and predicts average expenditures as or more accurately for most of the validation subgroups we considered. It appears to be the best single model empirically. However, for certain subgroups -- the elderly

[2] See Section 3.5 for a discussion of the limitations of the MCBS income variable.

receiving help with activities of daily living, for example, -- it doesn't appear to predict expenditures as accurately as certain of the survey models. No model predicts uniformly well for all groups. Thus, which model is preferred depends in part on what relative weight policymakers put on "getting payment right" for different subgroups of Medicare beneficiaries.

Practical and administrative considerations are also important in evaluating alternative claims and survey adjusters. The DCG/HCC model requires ambulatory diagnoses and hence encounter data systems, which are expensive and time-consuming to develop, although useful for a variety of purposes once implemented. The PIPDCG model requires only principal inpatient diagnoses, which are typically much less expensive to obtain than ambulatory diagnoses, of greater clinical validity, and easier to audit. Claims adjusters are sensitive to intentional and unintentional variations in diagnostic coding (e.g., "upcoding"). The PIPDCG model is much less sensitive to the completeness of diagnostic coding than the DCG/HCC model, but it is sensitive to which inpatient diagnoses are "principal" versus secondary. Moreover, a beneficiary must be hospitalized for his or her diagnosis to be taken into account in the PIPDCG model. This penalizes efficient health plans that avoid hospitalizations and sets up possibly inappropriate incentives for hospitalization.

Surveys have lower startup costs and are available more immediately than claims diagnoses, but are expensive and burdensome to conduct on an ongoing basis.[3] They

[3] The marginal cost of using surveys for risk adjustment may be low if the same survey instrument (e.g., the SF-36) that is used for outcomes assessment can also be used for risk adjustment (perhaps with a few added questions).

suffer from nonresponse, and biased and inaccurate responses (e.g., what does self-rated health mean from someone with dementia?). Providers may be able to influence survey responses (e.g., by "prescribing" disability) and beneficiaries may respond strategically once they realize that provider reimbursement depends on their survey answers. Survey responses may deviate from "objective" criteria along socio-demographic or regional lines, and are difficult to audit or verify. Either claims or survey adjusters would need to be recalibrated to adjust for behavioral changes in diagnostic coding or survey responses if either were implemented for payment.

Adding survey variables to a claims-based model such as the DCG/HCC increases overall explanatory power and improves predictions for key subgroups such as the elderly receiving help and dual Medicaid/Medicare eligibles. But a combined model requires obtaining both survey and encounter data, which might be prohibitively expensive. Also, overpredictions for certain diagnoses (osteoporosis, hip fracture, dementia) are increased. Combined models warrant more research as a means of combining diagnoses from claims with severity/disability information (subjective health, functional status) from survey responses into a single powerful model.

Substantial redundancy exists among the various survey adjusters. Their combined explanatory power is much less than the sum of their individual explanatory power. Nevertheless, independent dimensions of health status are measured by the different survey variables. A multi-dimensional survey model like our preferred survey model (see Chapter 6) or the SF-36 simulation is necessary to predict expenditures well across the

range of subgroups. The disadvantage of multi-dimensional survey models is that survey instruments must be longer, increasing survey expense and respondent burden, and lowering response rates.

Although multiple domains of health status need to be surveyed, redundancy implies that some pruning of questions based on other criteria is possible and desirable. Other desirable characteristics for risk adjusters include resistance to manipulation by providers or beneficiaries, objectivity, reliability, parsimony, and face validity. In our opinion, certain survey variables rank higher on these criteria than others. We would place chronic conditions (diabetes, heart disease), physical functioning ("can you walk two blocks?"), and activities of daily living ("do you have difficulty bathing") higher on this scale and social functioning ("has your health interfered with your social activities?"), self-rated health ("is your health excellent, good, fair, or poor?") and instrumental activities of daily living ("do you have difficulty preparing meals?") lower. Others might disagree with our assessment. More research and practical experience is needed on pertinent aspects of survey adjusters other than predictive power.

Even the best survey models don't predict accurately for groups defined on prior medical expenditures. Providers will be able to practice substantial risk selection against survey models by employing their knowledge of the medical care use of actual or potential enrollees. Although prior use and claims-diagnosis models worked well, our survey models also did not perform well for the disabled. At a minimum, parameters for the

elderly and the disabled appear to be different and require separate estimates. Perhaps totally different survey models need to be developed for the disabled.

A final, and very important point, is that more data are needed to obtain stable and reliable estimates of risk adjustment models. Although we believe that the combined year estimates of our preferred survey risk adjustment model (Chapter 6) are plausible and reasonably stable, they are less precise than would be ideal (as indicated by the relatively large standard errors of estimates). Not surprisingly, comparison of parameter estimates on single-year 1991/1992 versus 1992/1993 data shows substantial differences (Chapter 4). For example, "osteoporosis", which has a highly statistically significant coefficient of $1,312 in the 1991/1992 chronic conditions model has a negative and statistically insignificant coefficient of -$288 in the 1992/1993 model. "Very good" in the self-rated health status model has a statistically insignificant coefficient of $486 in 1991/1992 versus a highly significant coefficient of $1,003 in 1992/1993. More data will also increase the sensitivity of risk adjustment models by allowing the estimation of health status scales with more response levels (e.g., "a lot/some/a little/no difficulty" versus "some/no difficulty")

7.0 Limitations

This study has several significant limitations. We analyzed a particular set of survey models, albeit a wide range of this class of models. We analyzed only two claims-diagnosis-based models, the DCG/HCC and PIPDCG models, not, for example, the

Ambulatory Care Groups (ACG) model (Weiner *et al.*, 1996; Weiner *et al.*, 1991). Our results may not generalize beyond the particular survey and claims-based models we analyzed. Nor will our results necessarily generalize to other populations. For example, the differences between our results and those of Fowles *et al.*, 1996 may be attributable to that study's evaluation of the actual SF-36 survey scales and the Ambulatory Care Groups claims model on a mixed under-65 and aged population.

Because many high cost medical conditions that account for a large portion of expenditures are rare, large sample sizes, such as are available from claims files, are desirable for estimating and validating risk adjustment models. Sample sizes comparable to claims samples are not currently available for survey variables. Our MCBS results--both estimation and validation--may not fully generalize to other samples because of the MCBS's limited sample size. That is, our results are influenced to some extent by random error. Nevertheless, we believe that most of our qualitative findings will generalize to other and larger samples.

Another technical limitation is that we did not have a fully independent validation sample, which may tend to overstate the predictive power of all risk adjustment models. We did not include individuals in our sample for whom we did not have survey responses, either because a person did not respond to the MCBS at all, or because he or she did not answer a specific question. To the extent that survey nonrespondents are sicker on average than respondents, our results may somewhat overstate the predictive power of models (especially survey models) compared to a full set of responses.

8.0 Preferred Survey Model

Our analysis (Chapter 4) shows the importance of using large sample sizes to obtain stable estimates of the parameters of risk adjustment models. Subsequent to our primary analyses summarized above, we used three paired years of MCBS data to obtain more stable estimates of a survey risk adjustment model. Rather than reestimating all of our models, we focused on developing a single or preferred survey model that could be implemented from HCFA's Health of Seniors (HoS) survey. In consultation with HCFA staff, we selected MCBS survey variables appropriate for a payment model. Then, we coded MCBS questions to be comparable to items available on the HoS survey. For estimation, we pooled three pairs of MCBS years (1991/1992, 1992/1993, and 1993/1994). For each pair of years, health status information of the base (first) year is used to predict total annualized Medicare payments in the subsequent year. The pooled sample size is 32,335, including both elderly (sample size = 27,130) and the disabled (sample size = 5,205). Appropriate statistical corrections were made for the panel nature of the MCBS. Because all of the available data were used for estimation, the model was not validated.

Preliminary estimation revealed different parameters by aged/disabled subpopulations. Therefore, separate models were estimated for each group. Table ES-3 shows our final, preferred survey risk adjustment model for the aged. It is estimated on the pooled sample using variance components to account for longitudinal intra-person correlation. No R-squared is produced by the variance components routine, but the same

Table ES-3

Preferred Survey Risk Adjustment Model for Elderly

Dependent Variable: Annual Medicare Payments Per Beneficiary, 1992 Dollars

Variable	Coefficient	Standard Error	t Statistic
Intercept	790	204	3.87 ***
Age (65-74 omitted)			
75-84	661	166	3.99 ***
85+	1,134	268	4.23 ***
Male	658	164	4.02 ***
Self-Rated Health Status (excellent omitted)			
Poor	2,686	360	7.46 ***
Fair	958	259	3.70 ***
Good	547	218	2.51 **
Very good	311	214	1.45
Functional Status (no ADLs omitted)			
5-6 ADLs	1,687	386	4.37 ***
3-4 ADLs	860	328	2.62 ***
1-2 ADLs	510	210	2.43 **
Difficulty walking 2-3 blocks ("no" omitted)			
Unable/ a lot of difficulty	1,567	261	6.01 ***
Some/ a little difficulty	471	201	2.34 **
Difficulty lifting ("no" omitted)			
Unable/ a lot of difficulty	865	249	3.47 ***
Some/ a little difficulty	392	190	2.06 **
Chronic Conditions			
Heart Attack	1,169	244	4.79 ***
Angina	657	240	2.74 ***
Other Heart Conditions	571	183	3.11 ***
Cancer, except skin	457	201	2.27 **
Diabetes	1,476	217	6.80 ***
COPD	935	232	4.04 ***
R Squared	-		
Adjusted R Squared	-		
Observations	27,130		
F, Chi-Squared	359.2 ***		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor.
A variance components model was estimated using SAS PROC Mixed, which does not produce R-squares.
Coefficients and standard errors are in 1992 dollars.

*** Statistically significant at 1 percent level.
** Statistically significant at 5 percent level.
* Statistically significant at 10 percent level.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), 7(1993) and 10(1994).

risk\final\exectabs\tab ES-3\nd

model estimated using weighted least squares achieves an adjusted R-square of 4.6 percent. Coefficients and standard errors in Table ES-3 are in 1992 dollars. Although all variables in the final model are statistically significant, relative standard errors are still considerable for many coefficients. Reestimating the model on even larger sample sizes would be desirable to ensure parameter stability.

We were less successful in developing a final, preferred survey model for the under-age-65 disabled population. We recommend further study of survey models for this population before they are considered for use in payment.

1 Introduction

1.1 Background

Although it is a goal of the Medicare program to enroll more of its beneficiaries in managed care programs, currently only about 13 percent are enrolled in Health Maintenance Organizations (HMOs) or Competitive Medical Plans (HCFA Statistics - Table 6, Page 8). Greater managed care enrollment has the potential to reduce the growth in Medicare expenditures and improve the quality of care Medicare beneficiaries receive, and is consistent with private sector trends. Medicare's managed care shortfall has been attributed in part to the inadequacy of its current payment formula for HMOs--the Adjusted Average Per Capita Cost (AAPCC)--in accounting for expenditure differences among beneficiaries. The AAPCC considers only sociodemographic factors (age. sex, welfare status, institutional status, and employment status), location (county of residence), and reason for Medicare eligibility (aged or disabled). Many studies have shown that the AAPCC factors inadequately predict medical expenditures, creating inequities among HMOs that enroll healthier or sicker beneficiaries, and also large financial incentives for HMOs to try to attract healthier beneficiaries (Ellis *et al*., 1996). Numerous proposals have been advanced to incorporate additional factors into the AAPCC, such as health status, prior medical care use, diagnoses, and medical risk factors (Ellis *et al*., 1996). The additional factors that could be added to the AAPCC generally are

available from two sources: surveys of beneficiaries or medical claims data. If Medicare adopts a revised capitated payment methodology, it is likely to incorporate factors collected from either surveys or claims, or both. It is thus of interest to examine the merits of alternative survey and claims-based risk adjusters for the Medicare population.

The largest survey of the Medicare population currently available is the Medicare Current Beneficiary Survey (MCBS). This study employs three years (1991, 1992, and 1993) of the MCBS and associated Medicare claims data to evaluate alternative survey and claims-based risk adjusters on a common sample. With the requirement that Medicare claims for ambulatory patients contain diagnostic codes, there have recently been substantial innovations in claims-based risk adjusters (Ellis *et al.*, 1996; Weiner *et al.*, 1996; Kronick *et al.*, 1996). One of the latest generation of claims-based adjusters is included in our evaluation.

Both survey adjusters and claims-based adjusters have been extensively studied in the past, and continue to be a subject of intensive current research (see the references in Ellis *et al.*, 1996). But very few studies have compared both survey and claims-based measures. For example, Gruenberg *et al.* (1996) used the first two years of the MCBS to analyze survey risk adjusters. But that study did not include any claims-based measures, was limited to the elderly, noninstitutionalized Medicare population, and evaluated models using a different methodology than this study. Hornbrook and Goodman (1995, 1996) assessed the RAND-36 Health Survey and self-reports of chronic disease in a predominantly under-age-65 population enrolled in a large prepaid group practice HMO in the Pacific Northwest, but did not consider claims-based measures. Ellis *et al.* (1996), Weiner *et al.* (1996), and Kronick *et al.* (1996)

all analyzed only claims-based measures. Fowles *et al.* (1994; 1996) is the study most closely related to this one in that it evaluated both survey and claims-based risk adjusters. But the specific survey and claims models studied were different, and the population studied was very different--a predominantly younger, employed sample of enrollees in a Minnesota HMO versus a nationally representative Medicare elderly and disabled sample in our study.

1.2 Overview of the Report

This report consists of seven chapters and three appendices. Chapter 2 describes our data source, the Medicare Current Beneficiary Survey (MCBS), file and variable construction, and sample selection. Chapter 3 discusses methodological considerations, including sample weighting, complex survey design, functional form for estimation, missing data, and outliers. Chapter 4 presents specifications and estimates of alternative risk adjustment models, including survey-based, claims-based, and combined claims/survey models. The validation analysis of the Chapter 4 models is presented in Chapter 5, including validated predictive power of models for individuals and for nonrandom groups. Chapter 6 presents a final, preferred survey model estimated using pooled data from three pairs of MCBS years. This model could be implemented from responses to the Health of Seniors survey. Chapter 7 presents conclusions. Appendix A analyzes some additional MCBS disability measures not considered in Chapters 4 and 5, such as social and physical functioning. In addition, some of the previously included variables, such as Activities of Daily Living, are studied in a more disaggregated form. Appendix B explains how we simulated four SF-36-like health status

scales. Appendix C discusses required survey sample sizes to achieve specified levels of accuracy in survey-based risk-adjusted payment to capitated health plans.

2 Data and File Construction

2.1 Description of MCBS

The Medicare Current Beneficiary Survey (MCBS) is an ongoing, multi-purpose survey of a representative sample of the Medicare population. Survey data are linked to Medicare claims and other administrative data, making it one of the richest data bases available for research on the aged and disabled populations. This link enables researchers to analyze the utilization experience of respondents in terms of both self-reported data and provider claims. It contains self-reported information on health status, physical functioning, chronic conditions, risk behavior (e.g., smoking), supplemental insurance coverage, income, and other demographic information that may explain variations in utilization within the Medicare population. Claims data provide information on inpatient and outpatient utilization of Medicare-covered services.

Until recently, MCBS was a population-based survey that employed a panel design.[1] Computer-assisted personal interviews (CAPI) were conducted with a cohort of beneficiaries beginning in late 1991. This study used data from Rounds 1, 4, and 7 of the MCBS data which include information for 1991, 1992, and 1993. Data were collected from 12,674 persons in Round 1 (September - December, 1991). Approximately one year later, during

[1] After Round 10, a rotating sample design, similar to that used in the Current Population Survey, was adopted.

Round 4, 10,388 of these persons completed their follow-up interview. An additional 1,995 individuals were interviewed in Round 4 to account for attrition of the sample (due to death, relocation, and non-response). Round 7 interviews were completed in Fall 1993 with 10,936 individuals who participated in the earlier rounds as well as 1,927 sample replacements. Response rates for all three rounds were between 87 and 94 percent.

The data available for each round are organized by the primary sections of the survey: (1) identification, (2) health status and functioning, (3) access to care, (4) health insurance, and (5) enumeration (household and work information). For each of these sections, HCFA's Office of the Actuary (OACT) has created a "Record Information Code" (RIC) file where RIC 2 represents health status and functioning, *et cetera*. For respondents domiciled in facilities, there are two additional RIC files: RIC 6 on facility residence history and RIC 7 on facility identification. In addition to information gathered by the survey, OACT provides an additional file, RIC A, that contains information from Medicare's master eligibility and claims files. Among the more important items in the RIC A file are the data elements that indicate the expenditures (reimbursements) Medicare made on behalf of beneficiaries.

2.2 File Construction and Sample Restrictions

To construct our basic analytic file, data for Round 1 respondents were selected from the RIC files. Although data were obtained from nearly all of the files, most of the data came from the RIC A file (e.g., eligibility and date of death) and the RIC 2 file (almost all data

elements). Of the 12,677 Round 1 respondents, data on all but the following beneficiaries were obtained:

- three individuals interviewed in error;
- 120 individuals who died before January 1, 1992;
- 174 individuals who lived in Puerto Rico, U.S. territories, or otherwise outside of the United States; and
- 76 individuals who were ESRD entitled.

To the Round 1 data of the remaining 12,304 respondents, Round 4 RIC A expenditure data was added to the analytic file. Round 4 public use files do not contain information on attriters from Round 10. To obtain Round 4 expenditure data (based on Medicare claims) for the attriters, a special data request was submitted. For the attriters, OACT supplied us (1) an administrative record file that has the same format as the RIC A file and (2) with individual NCH claims.

After merging Round 4 expenditure data on to the analytic file, the sample was further restricted by dropping individuals – other than those who died during 1992 – who were not eligible for both parts of Medicare for all of 1992. This resulted in a sample size of 11,842 for the analytic file. Respondents who were enrolled at any time during 1991 or 1992 in a managed-care organization were excluded from the regressions because they would either have no Medicare claims or the Medicare claims would not capture any utilization provided by the managed-care organization. As described in Section 3.4, other restrictions reduced the number of observations available for analysis.

Parallel methods were used to construct the validation file that contains beneficiary characteristics during 1992 and expenditures during 1993. As before, expenditure data for attriters between 1992 and 1993 were obtained from special files constructed by OACT.

2.3 Expenditures

The objective of our analysis is to find a set of factors to predict 1992 Medicare expenditures for Medicare beneficiaries. The total Medicare payment variable was constructed by summing total Part A and Part B Medicare payments, variables that are available on the RIC A file. Enrollees who die account for a disproportionate share of the total reimbursements for Medicare enrollees, and thus should not be dropped or ignored when developing a payment classification system. Current payment policy is for HMOs to be reimbursed for each month a Medicare beneficiary is alive and enrolled in their plan. Hence, we have calculated payments for beneficiaries who die according to the number of months that they are eligible.

To develop the correct average payments for all beneficiaries in a payment class, including those who die, we use a process of weighting observations described in Ellis and Ash (1995). First, total payments are annualized by dividing by the fraction of the year each beneficiary is alive. For example, if a person is alive for six months and generates $6,000 of reimbursements, then the annualized payment for the person is $12,000. If this annualized amount were simply entered into regressions and calculations of means, then this would overstate the cost of such a person. Since sample weights are already present in the MCBS,

we adjusted the MCBS sample weights by multiplying them by the fraction of the year that the person is eligible for coverage. Hence, the MCBS weight for the person in the above example is multiplied by the fraction of the year eligible, which is 0.5. This process of annualizing and re-weighting observations results in unbiased estimates of the average and total payments for a group in which individuals are eligible for different fractions of the year.

Distributions of 1992 and 1993 Medicare program payments for beneficiaries in the analytic and validation samples are presented in Table 2-1. The two left-most columns have the distributions of annualized and weighted Medicare payments for 1992 and 1993. The two right-most columns have the distributions of Medicare payments for 1992 and 1993 that are neither annualized nor weighted. The effect of the annualization of payments is readily apparent from the much larger five highest values for annualized values than for the unannualized values.

Medicare payments per beneficiary were about eight percent higher in 1993 than in 1992 (using annualized and weighted payments). On an annualized payment basis, the values of the two most costly beneficiaries in 1992 ($2,019,504 and $873,312), however, are much higher than the two most costly beneficiaries in 1993 ($320,052 and $290,755). Because of the small sample sizes, regression model specification tests need to take into account the possible effect of the inclusion/exclusion of the highest payments during 1992. This will be discussed further in Section 3.4.

Table 2-1

Distribution of 1992 and 1993 Medicare Program Payments

	Annualized Total Medicare Payments (Weighted)[1]		Total Medicare Payments (Unweighted)	
	1992 N=10,893	1993 N=10,532	1992 N=10,893	1993 N=10,532
Mean	3,795	4,097	3,864	4,100
Standard Error	115	109	89	99
Quantiles				
100%	2,019,504	320,052	185,181	283,647
99%	61,403	63,875	44,424	43,606
95%	23,523	24,292	19,178	20,647
90%	13,283	13,485	11,625	12,013
75%	3,226	3,438	3,128	3,366
50%	519	600	510	593
25%	94	126	92	124
10%	0	0	0	0
5%	0	0	0	0
1%	0	0	0	0
0%	0	0	0	0
5 Highest Values				
	280,176	270,245	114,385	140,154
	320,388	280,415	140,322	146,932
	365,112	283,647	168,292	154,436
	873,312	290,755	168,392	257,047
	2,019,504	320,052	185,181	283,647

[1] Payments are annualized and weighted by the inverse of the annualization factor multiplied by the MCBS sampling weight.

SOURCE: MCBS Round 4 for 1992 payments and Round 7 for 1993 payments.

3 Methodological Considerations

3.1 Complex Sample Design

The MCBS has a complex rather than simple sample design. It is both stratified and has a clustered sample feature. We explicitly control for stratification in all analyses by weighting by the MCBS sampling weights. But we do not adjust for clustering, which requires special, difficult-to-use software such as SUDAAN.[1] While means and regression coefficients are not affected by cluster sampling, their standard errors are. Unless explicitly controlled for, cluster sampling usually results in higher standard errors.

Comparisons between weighted least squares (WLS) regressions (using SAS GLM), which control only for stratification, and SUDAAN regressions, which control for both stratification and clustering, were performed. The difference shows the effect of clustering on standard errors. The standard errors in the SUDAAN regressions are generally smaller than in the WLS regressions (see the example in Table 3-1). The statistical significance of some variables is affected. These results are surprising, since clustering is generally expected to result in larger standard errors. However, consultation with RTI staff who maintain SUDAAN indicate that it is possible that SUDAAN will produce smaller standard errors than WLS under certain circumstances. Since the small MCBS sample size affects the stability of

[1] SUDAAN is a specialized software package that adjusts for complex sample designs. It is maintained by the Research Triangle Institute (RTI).

Table 3-1

Comparison Between Weighted Least Squares (WLS) and SUDAAN Regressions: 1992 Medicare Payment, Regressed on Age, Gender, Self-Rate Health Status, Functional Status, and Selected Chronic Conditions

	WLS			SUDAAN		
Variable	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic
Intercept	1,250	326	3.83 ***	1,250	220	5.69 ***
Age (65-74 omitted)						
0-64	-1,220	452	-2.70 ***	-1220	359	-3.40 ***
75-84	513	263	1.95 *	512	190	2.69 ***
85+	1,194	420	2.84 ***	1194	349	3.42 ***
Male	238	242	0.98	237	205	1.16
Self-Rated Health Status (Excellent omitted)						
poor	2,364	527	4.48 ***	2364	582	4.06 ***
fair	957	411	2.33 **	956	307	3.11 ***
good	444	363	1.22	444	235	1.89 *
very good	194	372	0.52	193	217	0.89
Functional Status[a] (None omitted)						
5-6 ADLs	2,934	570	6.16 ***	2933	888	4.31 ***
3-4 ADLs	1,389	459	3.02 ***	1389	357	3.89 ***
1-2 ADLs	1,242	304	4.08 ***	1241	281	4.42 ***
IADLs only	117	487	0.24	117	306	0.38
Chronic Conditions						
Arteriosclerosis	583	354	1.65 *	583	367	1.59
Heart Attack	1,397	359	3.89 ***	1396	418	3.34 ***
Other Heart Conditions	615	283	2.17 **	615	221	2.78 ***
High Cost Cancer	1,254	687	1.82 *	1253	673	1.86 *
Diabetes	1,293	333	3.88 ***	1292	354	3.65 ***
Osteoporosis	863	455	1.89 *	862	478	1.81 *
Parkinson's Disease	1,566	926	1.69 *	1565	1,075	1.46
COPD	984	351	2.80 ***	983	337	2.92 ***
Partial Paralysis	433	453	0.96	433	453	0.96
Amputation of arm/leg	3,295	1,021	3.23 ***	3294	1,415	2.33 **
Lost urine > once per week	741	414	1.79 *	740	385	1.92 *
R Squared	0.0331					
Adjusted R Squared	0.0311					
Observations	10,893					
F	16.18 ***					

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor (see text).

[a] Definition differs from that used in subsequent chapters.

*** Statistically significant at 1 percent level.

** Statistically significant at 5 percent level.

* Statistically significant at 10 percent level.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991) and 4(1992).

the regression coefficient estimates, a conservative approach was adopted in which WLS regressions were used. The WLS regression estimates presented in this report use MCBS sampling weights adjusted for partial year Medicare enrollment (see Section 2.3).

3.2 Weighted Least Squares (WLS) versus Two-Part Model

Since a non-trivial proportion (10-15%) of Medicare beneficiaries have zero payments in a given year, we estimated two-part multivariate models (*Duan et al.*, 1983) of risk adjustment to compare with models based on WLS. Two-part models are also used to account for the skewness in expenditure for those with expenditures. Our analysis led to the conclusion that two-part models were not superior to the WLS models. Furthermore, WLS models are simpler and easy to interpret. In this section, we discuss one specification which compared WLS and two-part models.

Two-part models estimate the probability of any spending and the level of non-zero spending using separate equations. We used a probit model to estimate the likelihood of spending and an WLS model to explain the log of spending for those with positive spending. Spending was predicted for the entire sample using the predicted probability of spending and the level of spending after retransforming from the log scale. The model used various age categories separately for males and females, as well as Medicaid and institutionalization

status. Furthermore, we used functional limitation variables (ADL and IADL scores)[2] to control for health status.

We estimated 1992 spending using 1991 characteristics as regressors. Both WLS and two-part models had a similar R-square (2.1 percent).[3] Coefficients of marginal effects (Table 3-2) were also similar with the exception of the institutional status variable, which is imprecisely estimated. We also estimated WLS and two-part models using 1993 spending and 1992 characteristics for a similar specification. The R-square from the WLS model was considerably higher (2.4 percent) than the R-square in the two-part model (1.3 percent). Thus, WLS models performed better than (or at least as well as) the two-part models in estimation. Table 3-3 summarizes R-squares.

We also performed model validation using WLS and two-part models. Using estimates from the 1992 spending model, we predicted 1993 spending and compared actual and predicted 1993 spending for model validation. For the validation models, we calculated R-squares based on the difference between observed and predicted expenditures. The validation R-square from the two-part model was only slightly higher than the WLS R-square

---

[2] These models use an early definition of the ADL/IADL scale which is slightly different, and thus are not directly comparable to the revised estimates presented later in the report. Also, the model differs from the one shown in Table 3-1. The latter also includes self-rated health and chronic conditions.

[3] The R-square for the two-part model was calculated using the difference between actual and predicted spending where predicted spending was the product of the predicted probability of spending and the predicted level of spending. Data were retransformed from a log scale using the non-parametric smearing estimate (Duan et al., 1983) adjusted for age and sex.

Table 3-2

Weighted Least Squares (WLS) and Two-Part Estimated Marginal Effects with Functional Status

	1991 Characteristics/ 1992 Payments		1992 Characteristics/ 1993 Payments	
	Weighted OLS	Two-Part Model	Weighted OLS	Two-Part Model
Intercept	1,950.3		2,150.2	
	(300)		(322)	
Female, Age (65-69 omitted)				
0-64	-507.4	-522.4	-1,037.3	-843.4
	(536)		(716)	
70-74	120.5	145.7	192.9	-63.2
	(469)		(430)	
75-79	611.1	646.0	782.7	853.5
	(463)		(459)	
80-84	868.0	883.3	893.9	838.3
	(459)		(507)	
>84	1,415.6	1,194.6	1,498.8	1,157.3
	(482)		(549)	
Male, Age				
0-64	-875.7	-596.8	-83.4	34.6
	(464)		(601)	
65-69	578.0	487.6	486.8	748.8
	(437)		(469)	
70-74	870.0	1,119.4	1,146.5	1,773.6
	(517)		(465)	
75-79	982.1	1,109.9	2,179.2	2,969.8
	(525)		(523)	
80-84	1,249.9	1,613.2	2,586.3	2,767.6
	(556)		(629)	
>84	1,386.6	1,646.4	2,017.3	2,033.6
	(634)		(771)	
Medicaid Status	806.9	1,185.8	1,230.2	2,052.9
	(319)		(388)	
Institutional Status	-3,087.3	-1,069.3	-1,893.3	-590.6
	(531)		(638)	
Functional Status[a] (None omitted)				
5-6 ADLs	6,229.6	6,282.8	4,757.3	7,814.6
	(505)		(546)	
3-4 ADLs	3,496.0	4,336.1	3,699.4	4,981.2
	(413)		(423)	
1-2 ADLs	2,364.9	2,045.6	1,988.2	2,456.2
	(283)		(287)	
IADLs only	615.6	919.0	363.6	416.9
	(422)		(486)	
R-square	2.1%	2.1%	2.4%	1.3%

NOTE: Standard errors are in parentheses.

[a] Definition differs from that used in subsequent chapters.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), and 7(1993) for expenditures.

Table 3-3

Estimation and Validation R-Squares From WLS and Two-Part Models[a,b]

	Estimation		Validation
Model	1991-92	1992-93	1992-93
WLS	2.1 %	2.4 %	2.2 %
Two-Part	2.1	1.3	2.3

[a] AAPCC-like model with ADL scores (Table 3-2).

[b] Using weighted mean of spending and age-sex smearing in the two-part model.

(2.3 percent versus 2.2 percent).[4] We conclude that WLS models are preferable since they have about the same predictive power as the more complex, two-part models, but are simpler.

3.3 Sample for Analysis

We decided to analyze the full MCBS Medicare sample, subject only to the restrictions enumerated in Section 2.2.[5] The primary goal of our analysis is an exploratory investigation of the properties of survey and claims-based health status risk adjusters. The most information is gained if all populations are included in our estimation and validation analyses. A second stage of analysis, which is beyond the scope of this project, would be to develop specific models tailored to particular populations. Another factor in our decision to analyze the full sample is to maximize the rather small MCBS sample sizes, which increases the statistical power of our analyses. Finally, risk adjustment models that are relevant for any Medicare beneficiary enrolled by a health plan have the most general practical applicability.

Nevertheless, it is useful in the analysis to be aware of the different characteristics of important Medicare subpopulations. At least three important subsamples of the Medicare population that have distinct sociodemographic and health status characteristics can be distinguished. They are:

---

[4] In Table 3-3, it is surprising that the validation R-square exceeds the estimation R-square for WLS. This seems to be due to differences in the distribution of expenditures in 1993 versus 1992, i.e., fewer expenditure outliers in 1993. As will be seen in Chapter 4, R-squares, especially for certain models, tend to be higher in 1993 than 1992.

[5] For example, End Stage Renal Disease enrollees were excluded from our analysis because their care is reimbursed under a separate payment system by Medicare.

- the non-institutionalized elderly;
- the non-institutionalized disabled; and
- the institutionalized.[6]

Characteristics of each of these three subpopulations, and of the full sample, are shown in Table 3-4.[7]

The non-institutionalized elderly are the dominant Medicare group, constituting 78 percent of the sample observations. The non-institutionalized disabled account for 15 percent of the sample observations, or about 9 percent of the sample when weighted to adjust for oversampling. The institutionalized are the smallest group, accounting for only 7 percent of the sample observations. The disabled are slightly less expensive than the elderly, and the institutionalized are considerably more expensive than the non-institutionalized. The institutionalized also have by far the most expensive diagnoses, as measured by the Hierarchical Coexisting Conditions (HCC) score, an expected relative expenditure based on claims diagnoses (see Chapter 4 for further discussion of the DCG/HCC model).

The elderly, of course, are age 65 or over, while the disabled are under age 65. The institutionalized population is heavily weighted towards the oldest age group, age 85 and

[6] The "institutionalized" are those in long-term care facilities. A shortened version of the community interview is given to these respondents. Respondents are followed in and out of institutions in the multiple rounds of the MCBS. The MCBS defines long-term care facilities as having three or more beds and providing long term care services throughout the facility or in a separately identifiable unit. Types of facilities currently participating in the survey include nursing homes, retirement homes, domiciliary or personal care facilities, distinct long-term units in a hospital complex, mental health facilities and centers, assisted and foster care homes, and institutions for the mentally retarded and developmentally disabled.

[7] The subpopulations identified in this section are not the same as the MCBS sampling strata.

Table 3-4

Sample Characteristics (Weighted)[1], Overall and by Subsample

Variable	Full Sample	Non-Institutionalized Elderly	Non-Institutionalized Disabled	Institutionalized (Elderly & Disabled)
Observations (unweighted)	10,893	8,526	1,622	745
1992 Expenditures	$3,795	$3,752	$3,583	$4,951
HCC Scores	0.997	0.964	0.935	1.701
Age				
0 - 64	8.7 %	n/a %	100.0 %	11.9 %
65-74	49.7	56.4	n/a	13.1
75-84	31.5	34.6	n/a	30.0
85+	10.1	9.1	n/a	45.0
Male	41.1	39.9	61.1	29.5
Medicaid	11.6	7.3	33.1	53.7
Self-Rated Health Status				
Poor	10.3	7.7	35.8	13.8
Fair	21.0	19.1	30.1	41.8
Good	29.7	30.5	20.1	32.2
Very Good	23.0	25.1	8.8	9.9
Excellent	15.9	17.6	5.2	2.4
Functional Status				
5-6 ADLs	7.0	3.6	10.3	63.2
3-4 ADLs	9.3	8.0	17.6	17.8
1-2 ADLs	24.9	24.4	33.8	19.1
IADLs only	14.2	13.8	26.7	0.0
none	44.7	50.3	11.5	0.0
Chronic Conditions				
Arteriosclerosis	13.4	12.7	11.0	29.8
Heart Attack	13.6	13.4	16.9	9.8
Angina	13.3	12.8	15.4	19.6
Other Heart Conditions	24.5	23.8	25.2	35.0
Hypertension	47.4	48.7	40.4	35.8
Stroke	9.7	8.6	12.7	24.8
High Cost Cancer	2.8	2.9	3.2	0.4
Low Cost Cancer	12.8	13.3	10.3	7.7
Skin Cancer	13.3	14.5	5.5	4.5
Diabetes	14.4	14.2	15.7	16.0
Rheumatoid Arthritis	10.0	9.7	15.2	6.1
Osteoarthritis	44.8	46.0	38.0	34.1
Osteoporosis	7.2	6.9	6.7	14.9
Mental Retardation	2.1	0.2	16.7	10.4
Alzheimer's	3.0	1.1	1.0	42.1
Mental Disorders	5.5	2.3	29.0	24.1
Hip Fracture	4.4	3.6	3.5	20.6
Parkinson's Disease	1.6	1.3	1.3	6.9
COPD[2]	12.7	12.1	19.9	12.4
Partial Paralysis	7.6	5.7	21.7	17.8
Amputation of arm/leg	1.3	1.1	2.0	4.1
Lost urine > once per week	10.4	7.9	11.2	55.3

[1] All statistics below the "observations" row are weighted by the product of the MCBS sampling weight and the fraction of the year eligible.

[2] Chronic obstructive pulmonary disease.

SOURCE: Round 1 (1991) MCBS for characteristics and Round 4 (1992) for expenditures.

over. Females are more prevalent than males for all populations with the exception of the disabled. Only 7 percent of the elderly are Medicaid recipients, but one-third of the disabled and half of the institutionalized are Medicaid recipients reflecting spend-down to qualify for Medicaid nursing home coverage.

A much smaller proportion of the disabled or institutionalized populations report themselves to be in "very good" or "excellent" health status, and correspondingly, more report "poor" or "fair" health status. The disabled also report more functional limitations than the elderly, with the institutionalized reporting the highest levels of limitation. All the institutionalized report at least one activity of daily living (ADL) limitation. Among the self-reported chronic conditions, stroke, osteoporosis, dementia (Alzheimer's), mental disorders, mental retardation, hip fracture, Parkinson's Disease, partial paralysis, amputation, and especially urinary incontinence are disproportionately common among the institutionalized compared to the entire sample. Among the noninstitutionalized disabled, mental disorders, mental retardation, rheumatoid arthritis, chronic obstructive pulmonary disease, and partial paralysis are disproportionately frequent.

3.4 Missing Data and the Effects of Alternative Treatments for Outlier Payments

This section presents an analysis of missing data in the MCBS and the effects of payment outliers on analytic results. General sample exclusion criteria are described in Section 2.2. The exclusion of beneficiaries that were enrolled in an HMO during 1992 from the analyses was mainly responsible for the reduction in sample size from 12,304 to 11,190.

The subsequent exclusion of 1991 HMO enrollees reduced the sample size to about 11,000. Item non-response to survey questions was another source of the loss of observations from the sample. Preliminary comparisons of the explanatory power of the models made it clear that the comparisons were being confounded by the varying number of observations used in each regression. The varying number of observations in the regressions is due to item non-response. The number of observations used in the regressions for this preliminary analysis ranged from 10,453 to 11,190 when the socio-economic factors were included in the payment model. With such small sample sizes, we could not determine whether the differences in the R-squares were due to model specification or to the effect of the inclusion or omission of observations with outlier payments. We thus set the sample used in each regression, for this sub-analysis, to the largest size that results when all sets of explanatory variables are included as regressors in a payment model, 10,332 observations.[8]

By fixing the number of observations for each regression to the same sample, the confounding effect of the inclusion or omission of observations with outlier payments was removed from the comparisons of payment models. Observations with outlier payments (expenditures), however, have other effects. In particular, the skewness of the distribution of Medicare payments results in a non-normal distribution, violating standard OLS assumptions. Concern about the effect of observations with outlier payments in 1992 thus led to considering alternative treatments (transformations) for the dependent variable: (1) unedited, annualized Medicare payments during 1992, (2) natural log of annualized Medicare

[8] No renormalization of weights because of varying sample sizes was done.

payments during 1992, (3) re-coding (top coding) annualized Medicare payments during 1992 greater than \$50,000 to \$50,000 and (4) top coding annualized Medicare payments during 1992 greater than \$25,000 to \$25,000. The last three transformations of the dependent variable make the distribution of 1992 Medicare payment closer to a normal distribution. Top-coding payments is, in spirit, similar to risk pooling by HMOs participating in Medicare risk-based contracts, or risk sharing with the federal government.

The effect of transformations of Medicare payments can be seen in R-squares of different payment models (Table 3-5). For each model, the R-squares for transformed payments are higher than for the unedited payments. Top-coding payments at \$50,000 and then at \$25,000 results in successively larger R-squares. In the AAPCC factors plus chronic conditions model, for instance, the R-square increases from .035 (unedited), to .047 (top-coded at \$50,000), and then to .058 (top-coded at \$25,000). The relative rankings of the models, however, is largely unaffected by top-coding. Since we are interested in estimating models predicting the full range of Medicare expenditures and because Medicare does not currently have an outlier policy for HMO risk contracts, top-coding of Medicare expenditures is not used in subsequent chapters in this report.

3.5 Preliminary Consideration of Risk Adjusters

During the early phase of this project, variables (factors) that might explain 1992 expenditures were classified into several categories: (1) AAPCC classifying variables, (2) other socio-economic, (3) health risk, (4) functional status, (5) chronic conditions, (6) self-

Table 3-5

R-Squares from Regressions of 1992 Medicare Reimubrsement on AAPCC Factors and Other Explanatory Variables

	Total 1992 Medicare Reimbursement			
		Top-Coded at		
Explanatory Variables	Unedited	$50,000	$25,000	Natural LOG
AAPCC factors only	0.007	0.011	0.017	0.033
AAPCC factors and:				
Chronic conditions	0.035	0.047	0.058	0.108
Self-rated health status	0.025	0.034	0.044	0.065
Functional status	0.029	0.040	0.049	0.059
Health risk factors	0.011	0.016	0.021	0.055
Physical characteristics	0.011	0.016	0.023	0.040
Socio-economic factors	0.008	0.013	0.018	0.036
AAPCC factors, chronic conditions, and:				
Self-rated health status	0.041	0.055	0.066	0.114
Functional status	0.045	0.059	0.071	0.112
Health risk factors	0.037	0.049	0.060	0.118
Physical characteristics	0.036	0.048	0.059	0.109
Socio-economic factors	0.036	0.048	0.059	0.112
All factors	0.051	0.066	0.079	0.132

NOTES:

1. Sample size set to smallest size = 10,332 for all regressions.
2. Expenditures were annualized for beneficiaries that died during 1992.
3. Weighting factor for each observation adjusted for those who died in 1992.
4. "Top-coding" and the natural log were performed on expenditures after annualization.
5. The R-squares in the "Natural Log" column are not directly comparable to the other R-squares because of the transformation of the dependent variable.

SOURCE: 1991 and 1992 Medicare Current Beneficiary Survey.

rated health status, (7) physical characteristics, and (8) prior utilization. Comparisons of the explanatory power of each category were performed. Variables in the socio-economic, health risk, and physical characteristics categories were omitted from subsequent analyses based on the preliminary analysis presented in this section. We present our more detailed analysis of the remaining survey risk adjusters in Chapters 4 and 5.

The types of variables constructed for each omitted category and their regression results are briefly described in this section.

Socio-Economic Factors: Three variables representing socio-economic factors were constructed: (1) education, (2) income, and (3) living status. A review of the income variable found that an unrealistically large proportion of respondents indicated that their family income was less than $5,000 per year. Between OASDI and Supplemental Security Income (SSI) payments, the typical poor beneficiary should be receiving over $5,000 per annum.[9] While many SSI eligibles do not apply for and receive payments, the family income of elderly beneficiaries in the MCBS is much lower than the family income of elderly persons in the Current Population Survey (the source for many official income figures). This suggests that income may be under-reported in the MCBS as it is in many surveys. There also might be a reporting problem for beneficiaries residing in facilities since they are disproportionately represented in the lowest income brackets. Specifically, their SSI payments might be going directly to the facility (if it is Medicaid certified) and thus not being counted as income by the individual responding to the MCBS interviewer. The *living status* variable incorporates

[9] he average SSI payment was $286 per month in 1991 (1992 "Green Book"). The average OASDI payment for spouses of disabled workers was $153 per month in 1991–the lowest value in Table 1 in the Green Book was chosen. The sum of these two payments times twelve is $5,268.

information on marital status as well as whether the subject person lives with other individuals.

Miscellaneous Health Risk Factors: Three sets of variables representing miscellaneous health risk were created: (1) smoking, (2) obesity, and (3) vaccinations. The smoking variable distinguishes between individuals that have never smoked, individuals that currently smoke, and those who no longer smoke. The obesity variable was constructed by dividing body weight by height. The vaccination set is represented by two variables: (1) ever having had a shot for pneumonia and (2) having had a flu shot during the preceding winter.

Physical Characteristics: Two sets of variables representing physical characteristics were constructed: (1) eye/sight problems and (2) ear/hearing problems. Three eye problem variables were created that are recodes of the three MCBS variables concerning the functioning and status of eyes. Similarly, two "ear problem" variables were created that are recodes of the two MCBS ear/hearing problem variables.

Regression Results: The explanatory power of the socio-economic variables, health risk factors, and physical characteristics are shown in Table 3-5. In the top half of the table, where each of these sets of variables is added to the AAPCC factors, the R-squares are usually much lower than those for chronic conditions, self-rated health status, and functional status. Indeed, except for the regressions where the dependent variable (1992 Medicare payments) is in natural log form, the R-squares for the socio-economic variables, health risk factors, and physical characteristics are less than half of those for chronic conditions, self-rated health status, and functional status. The incremental increase in R-square of the socio-

economic variables, health risk factors, and physical characteristics over the AAPCC-only model is very slight, ranging from .001 to .006 except for the natural log where the incremental increase in R-square ranges from .003 to .022 (health risk factors).

In the bottom half of Table 3-5, the R-squares are presented for the socio-economic variables, health risk factors, and physical characteristics in conjunction with AAPCC factors plus chronic conditions. In relative terms, the socio-economic variables, health risk factors, and physical characteristics do not perform as poorly in the bottom half of the table as in the top half. Indeed, for the natural log of Medicare payments, the R-squares for the socio-economic variables, health risk factors, and physical characteristics are as high as for self-rated health status and functional status. The incremental increase in R-squares for the socio-economic variables, health risk factors, and physical characteristics over the AAPCC factors plus chronic conditions models, however, is generally low, ranging from .001 to .002 except for the natural log model where health risk factors increased the R-square by .01.[10]

These model comparisons lead us to exclude the socio-economic, health risk, and physical characteristics categories from subsequent analyses because of their low predictive power. In addition, regardless of explanatory power, the socio-economic variables would not have been retained in a payment model since it is unlikely that the federal government would ever pay on the education level of a beneficiary, for example. A consequence of this decision

---

[10] The relative increase in R-square due to the health risk factors, however, is not so different in the natural log model. The baseline R-square is higher in the log model because of the transformation of the dependent variable. This higher R-square does not necessarily indicate that the log model has more predictive power, since it is predicting a different dependent variable.

to drop the aforementioned categories was to increase the 1991/1992 sample size to 10,893 for the remainder of this report.

4 Estimation of Alternative Risk Adjustment Models

4.1 Overview

We developed nine classes of risk adjustment models using the information available from the MCBS. They are:

1. demographic;
2. self-rated health status;
3. self-reported chronic conditions;
4. functional status;
5. comprehensive survey;
6. short form (SF)-36 simulation;
7. claims diagnoses;
8. claims diagnoses plus survey variables; and
9. prior use.

This chapter discusses the specification and estimation of these models. Model validation and evaluation is reported in Chapter 5. Each model was estimated using 1991 (Round 1) characteristics to predict 1992 Medicare program expenditures. Expenditures are annualized and ordinary least squares regression estimates are weighted by the product of MCBS sampling weights and the fraction of months each sample person was alive and Medicare eligible in 1992. The sample includes the elderly, disabled, and institutionalized, subject to the restrictions given in Chapter 2. Although the estimation coefficients and R-squares reported in this chapter are useful information, the true test of a model is the out-of-sample (i.e., a different sample) predictive power reported in Chapter 5.

We also report estimates of the risk adjustment models using 1992 characteristics to predict 1993 expenditures.[1] Stability of coefficient estimates is compared between 1991/1992 and 1992/1993.[2] The stability issue is important given the small sample sizes available on the MCBS. However, model selection for validation (e.g., the particular list of chronic conditions for the chronic condition model) was based strictly on the 1991/1992 estimates. This was necessary because our validation analysis uses 1992/1993 data, and estimation/model development must be separated from validation for a proper test of the models. Thus, the 1992/1993 estimates play no role in either model development/selection or in validation; they are reported simply for comparative purposes (i.c., stability analysis).[3]

All models include categorical variables for age and gender. The categorical variable for the 0 to 64 age category allows the intercept to shift for the disabled-entitled population, all of whom are under age 65. Institutional and Medicaid status were excluded from all models except the AAPCC-like demographic model. These variables are potentially endogenous (e.g., institutional status can be affected by provider activities) and subject to varying definitions. Their relationship to expenditures is not well understood. Controlling only for age and gender in addition to health status allows for a more easily interpretable evaluation of the risk adjustment models, free of any confounding effects of the two variables.

---

[1] The SF-36-like and comprehensive survey models were estimated only on 1991/1992 data.

[2] The latter set of estimates are not adjusted to 1992 dollars. Thus, the 1992/1993 coefficients will tend to be somewhat higher merely by the fact of higher average expenditures in 1993.

[3] The validation analysis (Chapter 5) uses 1992 health status information multiplied by 1991/92 coefficient estimates, to predict 1993 payments, which are then compared to actual 1993 payments. The 1992/93 estimates were derived by regressing 1993 payments against 1992 health status variables.

Institutionalization and Medicaid status are included in the analysis as validation groups (see Chapter 5).

Our basic strategy was to define a range of models representing the health status information available on the MCBS. Then, the risk adjustment properties of the health status measures are analyzed individually and in combination. Demographics are the standard that has been historically and currently used for risk adjustment. Demographic models provide a baseline that other models can be compared to.

A model is defined for each of the three major survey health status measures--self-rated health (also called "general" or "perceived" health status), self-reported chronic conditions, and functional status--so that the properties of each measure can be isolated. These measures are also combined into a comprehensive survey model to analyze their joint properties. We use standard scales of self-rated health and functional status. The MCBS collects a wealth of functional status variables from which a variety of disability measures can be constructed. Further exploration of alternative functional status models is reported in an appendix. We also developed from the MCBS a model that partially simulates the widely-used 'SF-36' health status scales (Ware and Sherbourne, 1993). This model incorporates self-rated health, elements of functional status, and social functioning.

We estimated two Diagnostic Cost Group, or DCG models (Ellis *et al.*, 1996) to provide a comparison to the survey models. The DCG/PIP, or PIPDCG model, relies on principal inpatient diagnoses from hospitalizations only. The DCG/HCC model uses all the diagnoses recorded on hospital and physician claims. We calculated PIPDCG and DCG/HCC

expenditure predictions based on Medicare claims for the MCBS sample. Unlike the other models, the DCG models were not estimated on the MCBS sample, which is too small to properly estimate the parameters of these models. Rather we use the estimates obtained from the much larger 2.5 percent Medicare sample (n = 680,000) used by Ellis *et al.* (1996). The more stable parameter estimates from this much larger sample give the DCG models some advantage over the MCBS-estimated models in the validation analysis (Chapter 5).

In addition to the DCG models themselves, we added survey measures to the DCG scores to form two additional models. These "combined" models allow us to evaluate the incremental contribution of survey variables to the DCG models. Finally, we developed a model based on prior use of medical services for comparison to the survey and claims-diagnosis models.

4.2 Alternative Models

We now turn to a specific discussion of the formulation and estimation of alternative risk adjustment models. The eight model types listed in Section 4.1 are discussed in the eight subsections that follow.

4.2.1 Demographic Models

Demographic factors are the only variables that are currently widely used for risk adjustment. This is because they are readily available in administrative records and many are

difficult for providers or beneficiaries to manipulate. We consider two demographic models:

(1) an age/gender model, the most basic risk adjustment model; and

(2) the factors included in Medicare's AAPCC, which add Medicaid enrollment and institutional status to age and gender.

Age/Gender Model. Table 4-1 shows estimates of a simple age/gender model. This represents the most basic risk adjustment model, a baseline for comparison to all other models. These risk adjustment variables, age and gender, are included in all other models. We collapsed age into four cells: 0-64, 65-74, 75-84, and over 85. A categorical age variable allows nonlinear relationships with cost, but further division tends to produce small sample sizes in the MCBS and unstable estimates. The 0-64 category is synonymous with Medicare beneficiaries who are entitled by reason of being disabled. The elderly are broken into three age categories. Gender is not interacted with age, but entered as additive to age, again to provide larger cell sizes. This specification assumes that the incremental effect of gender is equal at all ages.

The 1991/1992 coefficients show that the disabled are not significantly more expensive than the youngest elderly. Mean expenditures do rise with age among the elderly, with the oldest age group (85+) more than two thousand dollars more expensive than the youngest elderly (65-74). Males and females are about equally expensive. The percentage of expenditure variation explained (R-square) is a low 0.4 percent, which is probably accounted for by the few large expenditure outliers in the 1992 MCBS data.

Table 4-1

Regression of Medicare Payments On Age and Gender

	1991 Characteristics/ 1992 Payments			1992 Characteristics/ 1993 Payments		
Variable	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic
Intercept	3,164	193	16.42 ***	2,883	199	14.49 ***
Age (65-74 omitted)						
0-64	363	425	0.85	351	318	1.11
75-84	958	263	3.65 ***	1,650	263	6.28 ***
85+	2,402	400	6.00 ***	2,826	344	8.22 ***
Male	136	237	0.57	640	224	2.86 ***
R Squared	0.0037			0.0084		
Adjusted R Squared	0.0032			0.0079		
Observations	10,893			10,532		
F	10.24 ***			22.26 ***		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor.

*** Statistically significant at 1 percent level.

** Statistically significant at 5 percent level.

* Statistically significant at 10 percent level.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), and 7(1993).

For comparison, 1992/1993 coefficient estimates for the model are shown. The two sets of coefficient estimates are substantially different in absolute terms, differing by more than $500 in several instances. However, the variation is usually within the range indicated by the 1991/1992 standard errors, and so can be attributed to random small sample variations, rather than any fundamental shift in the relationship among age, gender, and cost. The gradient of cost with age is steeper in the 1992/1993 sample. Males are also significantly more expensive than females. The fewer 1993 expenditure outliers result in a higher R-square of 0.8 percent.

AAPCC-Like Model. Medicare's current payment model for HMOs, the Adjusted Average Per Capita Cost or AAPCC, uses Medicaid enrollment and institutional status in addition to age and gender. The AAPCC defines many interactive cells among these variables, with predefined "ratebook" payment weights for each cell. Payments are also adjusted for county of beneficiary residence. We did not attempt to replicate or simulate the full-blown AAPCC with the MCBS. Rather, we estimate a simple noninteractive, additive linear model using the AAPCC factors. No geographic adjustments are included. This model is presented in Table 4-2.

As compared with age/gender alone, Medicaid and institutional status increase the R-square from 0.4 percent to 0.5 percent (1991/1992) and from 0.8 percent to 1.1 percent (1992/1993). Virtually all the increase in explanatory power comes from Medicaid enrollment, which has a large (about $1,500) and statistically significant coefficient.

Table 4-2

Regression of Medicare Payments On Age, Gender, Medicaid Status, and Institutional Status ("AAPCC-Like")

	1991 Characteristics/ 1992 Payments			1992 Characteristics/ 1993 Payments		
Variable	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic
Intercept	3,033	196	15.48 ***	2,707	202	13.43 ***
Age (65-74 omitted)						
0-64	-40	439	-0.09	-217	335	-0.65
75-84	924	263	3.51 ***	1,568	263	5.95 ***
85+	2,260	416	5.44 ***	2,513	359	7.00 ***
Male	226	238	0.95	766	225	3.41 ***
Medicaid	1,399	388	3.6 ***	1,527	320	4.78 ***
Institutionalized	-190	584	-0.33	505	480	1.05
R Squared	0.0050			0.0110		
Adjusted R Squared	0.0045			0.0104		
Observations	10,893			10,532		
F	9.08 ***			19.56 ***		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor.

*** Statistically significant at 1 percent level.

** Statistically significant at 5 percent level.

* Statistically significant at 10 percent level.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), and 7(1993).

Institutionalization, on the other hand, has a small and insignificant relationship to costs. [4] Surprisingly, those who are institutionalized are not more expensive controlling for age, gender, and Medicaid status. Although nursing home residents are expected to be more frail and disabled, and thus more costly, there is probably a substitution between Medicare's acute care services and nursing services in an institution. Also, some reasons for institutionalization (e.g., dementia) do not necessarily imply higher acute care costs. These factors appear to balance out to produce no net effect of institutionalization on cost in a demographic model. The additional AAPCC factors reduce the age coefficients somewhat, but older age remains a highly significant predictor of future costs.

4.2.2 Self-Rated Health Status

The limited predictive power of demographic variables has led to a search for additional factors that can predict medical costs. Health status is an obvious candidate. Self-rated health status is probably the most widely-used measure of general health status. It has the advantages of parsimony and generality, but the disadvantages of subjectivity and lack of specificity.[5] It has been previously studied for risk adjustment, and has been widely used for other purposes, such as outcomes and quality of care assessment. The self-rated health status

[4] See Section 3.3 for the MCBS's definition of institutionalization.
[5] The subjectivity of this question is particularly an issue for the cognitively impaired, or in cases where a proxy respondent must assess health status.

question on the MCBS, which is standard, is

> In general, compared to other people your age, would you say that your health is excellent, very good, good, fair, or poor?

Note that health is assessed relative to other people your age, which means self-rated health status is a relative, not absolute, measure of health that should be largely orthogonal to age. Also, the question asks for health "in general", which should focus the respondent on more of a long-run or chronic concept of health stock than (acute) health "at this moment" or "in the last two weeks". Self-rated health should be relatively powerful at making distinctions among people in good health because it uses three of its five categories ("good", "very good", and "excellent") to differentiate good health states.

Self-rated health is an ordinal scale with five levels. This construction naturally suggests a regression specification of four binary variables for the different levels of health ("excellent" is omitted to avoid perfect collinearity with the regression constant term). This specification does not impose any a priori constraints on the coefficients of the different levels of health. Table 4-3 shows MCBS estimates of this specification.

Adding self-rated health status to age and gender raises the R-square substantially, from 0.4 percent to 1.9 percent in 1991/1992. With the exception of "very good," all the health categories have large and significant coefficients. The ranking of the coefficients is consistent with their scaling of health (i.e., those in poorer health cost more). Those in "poor" health cost, on average, $5,000 more than those in excellent health, ceteris paribus, a

Table 4-3

Regression of Medicare Payments On Age, Gender, and Self-Rated Health Status

	1991 Characteristics/ 1992 Payments			1992 Characteristics/ 1993 Payments		
Variable	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic
Intercept	1,708	323	5.28 ***	964	311	3.10 ***
Age (65-74 omitted)						
0-64	-1,132	439	-2.58 ***	-1,234	329	-3.74 ***
75-84	839	261	3.22 ***	1,370	261	5.26 ***
85+	2,138	398	5.37 ***	2,424	341	7.11 ***
Male	224	236	0.95	724	221	3.27 ***
Self-Rated Health Status (excellent omitted)						
Poor	5,197	470	11.05 ***	6,447	444	14.51 ***
Fair	2,702	383	7.05 ***	3,427	364	9.42 ***
Good	1,335	356	3.75 ***	2,297	336	6.83 ***
Very Good	486	373	1.30	1,003	348	2.88 ***
R Squared	0.0185			0.0323		
Adjusted R Squared	0.0178			0.0316		
Observations	10,893			10,523		
F	25.62 ***			43.89 ***		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor.

*** Statistically significant at 1 percent level.

** Statistically significant at 5 percent level.

* Statistically significant at 10 percent level.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), and 7(1993).

substantial difference. Adding self-rated health status turns the age 0-64 (disabled) coefficient negative.[6] This is because many disabled rate themselves in poor health. At a given level of self-rated health status, the disabled are less costly than the elderly. The elderly age coefficients, however, are not greatly affected by the addition of self-rated health status. This is consistent with the wording of the MCBS question, which asks respondents to rate their health relative to other people their own age.

The self-rated health status coefficients are somewhat unstable from the 1991/1992 to the 1992/1993 samples, although not outside the range that is consistent with random variation and outlier effects. The 1992/1993 coefficients are larger and more significant at all health levels. "Poor" health has a particularly larger dollar coefficient in 1992/1993, which could be the result of one or a few large 1993 spending outliers given the relatively small number of MCBS respondents in this category. But a few outliers do not explain the larger "across the board" strength of self-rated health status in 1992/1993. The percent of variation explained (R-square) shows the largest increase of any of the models from 1991/1992 to 1992/1993. The reasons for this are not entirely clear, but could be due to random small sample or outlier effects. The good fit of the self-rated health model to the 1992/1993 data gives it an advantage for the validation analysis, which is done on the 1992/1993 data (see Chapter 5).

[6] Note that the model never implies negative payments, because the negative disabled coefficient is added to the larger in absolute value positive intercept term. However, a disabled female reporting excellent health would induce a capitation payment of less than $600.

4.2.3 Self-Reported Chronic Conditions

Diagnostic information is another source of information to predict medical care costs. People with certain diagnoses, especially chronic ones, cost more now and in the future. Substantial research verifying this has been done with International Classification of Disease, Ninth Revision, diagnoses recorded on Medicare provider claims (e.g., Ellis *et al.*, 1996). Self-reported diagnoses, as on the MCBS, have the advantage of not requiring claims or encounter data. They have the disadvantages of being less complete since respondents can only be queried about a limited list on a survey, less specific since beneficiaries are less clinically knowledgeable than physicians and medical records coders, and less accurate since beneficiaries suffer memory lapses.[7]

The MCBS asks respondents about a list of chronic conditions. In 1991 (Round 1), the question is worded:

> Next, I'm going to read a list of medical conditions. Please tell me if a doctor has ever told you that you had any of these conditions.

In 1992 (Round 4), for Round 1 respondents the question is modified to ask if a doctor had told you <u>in the last year</u> if you had any of these conditions. New respondents are asked the Round 1 version of the question. To measure 1991 chronic conditions, we used the Round 1 version of the question ("have you ever had the condition"). To measure 1992 chronic conditions, we used a "yes" response to either the Round 1 or Round 4 version of the question ("have you had the condition ever or in the past year").

[7] The severity and the duration of illness (or conditions or diseases) are hard to measure through a survey or by use of claims.

Note that the two versions of the question may have different cost implications. If the person has seen a doctor who told him or her that he or she had the condition in the past year, the condition is likely to be more active and in need of treatment. People who had a condition (e.g., a heart attack) in the distant past and have now recovered are likely to be less costly than those who had an acute form of the condition in the last year. Note also that the "ever had" element of the MCBS question may pick up people who do not have the diagnosis recorded on a provider claim in the past year.

The MCBS question asks respondents to report a diagnosis only if a doctor has told them that they have it. This limits inaccuracies associated with self-diagnosis, but also requires medical utilization (a physician visit) for a diagnosis to be reported. However, most of the MCBS conditions are serious enough that health care utilization associated with them would be expected.

The definitions of several of the medical conditions require clarification. The MCBS asks whether the person has "cancer, malignancy, or tumor" for several body parts, including "other", which the respondent is then asked to specify. Based on our DCG work (Ellis *et al.*, 1996), we divided the cancer response into two categories "high cost" and "low cost" by affected body part. Lung, ovary, kidney, stomach, brain, throat, or head were considered high cost, all other cancers (e.g., colon, breast, prostate) low cost. Unfortunately, in its question, the MCBS explicitly confounds the lower-cost "benign or non-malignant tumors or growths" with cancer. We also included the response to the question "have you lost urine beyond control in the past 12 months (more than once per week)" in the list of conditions. Although

the MCBS did not include urinary incontinence in its condition list, and is not asked in reference to information received from a doctor, it seems to belong in the chronic conditions model.

Regressions of Medicare expenditures on age, gender, and the full MCBS list of medical conditions are shown in Table 4-4. Not surprisingly, given the small sample size of the MCBS and the long list of conditions, many of the condition coefficients are statistically insignificant and unstable between 1991/1992 and 1992/1993. The instability is least for the more common conditions, and greater for the rare conditions. Some conditions actually have negative coefficients, and the negative Alzheimer's coefficient is significant at the 10 percent level in both 1991/1992 and 1992/1993. We attribute this either to underservice to demented beneficiaries who cannot express their medical needs adequately, and/or a substitution of nursing home care for acute care for these people.

All conditions with negative or statistically insignificant coefficients in the "complete" 1991/1992 regression were eliminated to yield the "selected" self-reported conditions model reported in Table 4-5.[8] All the 1991/1992 coefficients in the selected conditions model are large and highly significant. But on the 1992/1993 sample, many of the coefficients are insignificant, and one (osteoporosis) is actually negative! This vividly illustrates the dangers of choosing variables for a model using a small sample. The "selected conditions" model, rather than the "full" model, was validated (Chapter 5).

[8] In future work, stepwise regression techniques could be used to account for collinearity among conditions when selecting from the full set for a reduced model.

Table 4-4

Regression of Medicare Payments On Age, Gender, and Chronic Conditions

	1991 Characteristics/ 1992 Payments			1992 Characteristics/ 1993 Payments		
Variable	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic
Intercept	1,736	263	6.60 ***	1,490	262	5.69 ***
Age (65-74 omitted)						
0-64	-164	479	-0.34	-45	368	-0.12
75-84	709	263	2.70 ***	1,279	264	4.85 ***
85+	1,926	415	4.64 ***	2,230	362	6.17 ***
Male	186	247	0.75	668	232	2.88 ***
Chronic Conditions						
Arteriosclerosis	777	365	2.13 **	-76	338	-0.23
Heart Attack	1,390	377	3.68 ***	1,240	355	3.49 ***
Angina	582	384	1.52	1,163	358	3.24 ***
Other Heart Conditions	826	286	2.89 ***	1,328	264	5.03 ***
Hypertension	41	236	0.18	67	225	0.30
Stroke	565	440	1.28	380	387	0.98
High Cost Cancer	1,343	691	1.94 *	325	606	0.54
Low Cost Cancer	253	347	0.73	146	309	0.47
Skin Cancer	-282	340	-0.83	-188	316	-0.59
Diabetes	1,697	333	5.10 ***	1,912	310	6.17 ***
Rheumatoid Arthritis	116	388	0.30	682	329	2.07 **
Osteoarthritis	92	236	0.39	27	228	0.12
Osteoporosis	1,244	459	2.71 ***	-386	393	-0.98
Mental Retardation	-336	879	-0.38	-1,299	599	-2.17 **
Alzheimer's	-1,332	714	-1.87 *	-935	559	-1.67 *
Mental Disorders	259	540	0.48	1,203	411	2.93 ***
Hip Fracture	397	573	0.69	84	493	0.17
Parkinson's Disease	2,521	926	2.72 ***	1,033	802	1.29
COPD	1,421	348	4.08 ***	1,273	319	4.00 ***
Partial Paralysis	858	493	1.74 *	376	433	0.87
Amputation of arm/leg	3,969	1,022	3.88 ***	1,555	937	1.66 *
Lost urine > once per week	1,832	402	4.55 ***	1,687	335	5.04 ***
R Squared	0.0262			0.0344		
Adjusted R Squared	0.0239			0.0320		
Observations	10,893			10,532		
F	11.24 ***			14.40 ***		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor.

*** Statistically significant at 1 percent level.
** Statistically significant at 5 percent level.
* Statistically significant at 10 percent level.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), and 7(1993).

Table 4-5

Regression of Medicare Payments On Age, Gender, and Selected Chronic Conditions

	1991 Characteristics/ 1992 Payments			1992 Characteristics/ 1993 Payments		
Variable	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic
Intercept	1,851	213	8.70 ***	1,637	217	7.54 ***
Age (65-74 omitted)						
0-64	-158	427	-0.37	81	320	0.25
75-84	693	261	2.65 ***	1,282	262	4.89 ***
85+	1,860	404	4.60 ***	2,212	351	6.29 ***
Male	127	242	0.53	620	228	2.72 ***
Chronic Conditions						
Arteriosclerosis	913	354	2.58 ***	235	325	0.72
Heart Attack	1,626	358	4.54 ***	1,660	335	4.96 ***
Other Heart Conditions	904	281	3.21 ***	1,523	258	5.89 ***
High Cost Cancer	1,402	689	2.03 **	377	604	0.62
Diabetes	1,774	329	5.39 ***	2,054	306	6.72 ***
Osteoporosis	1,312	453	2.90 ***	-288	386	-0.75
Parkinson's Disease	2,446	923	2.65 ***	1,070	799	1.34
COPD[a]	1,452	347	4.19 ***	1,383	317	4.36 ***
Partial Paralysis	1,174	444	2.64 ***	577	398	1.45
Amputation of arm/leg	3,989	1,021	3.91 ***	1,626	938	1.73 *
Lost urine > once per week	1,744	391	4.47 ***	1,643	323	5.09 ***
R Squared	0.0253			0.0316		
Adjusted R Squared	0.0240			0.0302		
Observations	10,893			10,532		
F	18.85 ***			22.85 ***		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor.

[a]COPD: Chronic obstructive pulmonary disease.

*** Statistically significant at 1 percent level.
** Statistically significant at 5 percent level.
* Statistically significant at 10 percent level.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), and 7(1993).

Nevertheless, the self-reported conditions model achieves a relatively high R-square in both years, better than either self-rated health status or functional status in 1991/1992 and second to self-rated health status in 1992/1993. The age 0-64 (disabled) coefficient does not turn as negative in the conditions model as in the self-rated health or functional status models. Given a mix of diagnoses, the disabled cost more nearly the same as the elderly.

4.2.4 Functional Status

Background. Functional status measures, like Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs) provide valuable information about health status beyond diagnostic information. First, impairments in ADLs and IADLs serve as severity measures for the impact of individual conditions or combinations of medical conditions on the individual. For example, a diagnosis or report of a cardiac condition does not differentiate between the individual whose cardiac status is so compromised that s/he cannot perform basic activities without difficulty and the individual whose cardiac condition is less advanced or severe. Second, the degree of impairment itself has implications for future health in relation to the resultant decreased activity and mobility. A person who has limited mobility, whether it is due to arthritis pain or cardiac function, is at risk of deconditioning, increased risk of cardiopulmonary problems including pneumonia, and increased risk of falling. A severely ADL compromised individual is also at risk of skin breakdown. IADL limitations also serve as measures of disease impact and as risk factors for inadequate nutrition and hydration.

Description. In this section, we discuss the basic elements of our approach to functional status measurement and the findings for the model we selected. We chose to use a scale instead of individual functional status questions. We also defined functional status limitations in terms of self-report of difficulty due to a health reason, as opposed to receiving help to perform an activity. To arrive at these modeling decisions, we evaluated both a priori considerations and the results obtained from running different models. We present detailed information about the functional status literature, the functional status measures in the MCBS, and the alternative models that we evaluated in Appendix A.

Elements of the Functional Status Model. As with all of our models, we included the three age categories 0-64, 75-84 and 85 and over, with 65-74 as the omitted category. We also included gender, with female gender as the omitted category.

We created a scale based on the number of Activities of Daily Living (ADLs) and the presence of Instrumental Activities of Daily Living (IADLs) but no ADLs. The scale is based on the premise that more ADL impairments indicate greater disability. The scale is:

- 5-6 ADL impairments;
- 3-4 ADL impairments;
- 1-2 ADL impairments;
- IADL impairments only; and
- No ADL or IADL impairment.

An activity was coded as impaired if the sample member or proxy reported either difficulty or inability to perform the activity due to a health reason. The ADLs included bathing, dressing, walking in the home, transferring, toileting and eating. The IADLs included light housework, heavy housework, meal preparation, shopping for personal items, use of the

telephone, and money management. The nursing home sample members were only asked about shopping for personal items, use of the telephone and money management. We imputed difficulty in light and heavy housework and meal preparation to the nursing home sample.

We chose to include the functional status data based on a scale rather than individual ADLs and IADLs to avoid overfitting the model to the MCBS given the sample size and to allow for the effects of varying combinations of variables. The MCBS sample is small and the total number of persons with any functional limitations or with specific limitations are too few to validate through split sample or repetitive random sampling techniques. Other researchers (Gruenberg, Kaganova and Hornbrook, 1996) reported instability of individual ADLs using a bootstrapping, or repetitive random sampling approach. Our analyses indicate instability of individual tasks as predictors between the 1991/92 models and the 1992/93 models. The observed instability may be related either to sample size or to an actual lack of reliability of the individual items as predictors. The Katz ADL scale, upon which the ADL variables in the MCBS are based, has been demonstrated to have good qualities as a scale, including high coefficients of reproducibility and scalability (Spector, 1996). We also found the frequencies and correlations among the ADL items in the MCBS to support the scale approach.

There are disadvantages to using a scale. The scale treats each ADL symmetrically. In addition, while the measure is parsimonious, it requires collecting information on all the ADLs and IADLS.

We chose to define functional impairment based on a report of difficulty or inability to perform for a health reason, rather than on reported receipt of help, for several reasons. First, our purpose is to evaluate risk adjusters for the general Medicare population, unlike payment methodologies for demonstrations targeted to the smaller segment of the population that is at risk of institutionalization. To this purpose, we believe the salient cut is between those who are healthy and those who are impaired in any degree, and to develop risk adjusters that are predictive for larger segments of the population. Close to 50 percent of the sample report difficulty with at least one ADL, while slightly under 25 percent report receipt of help with at least one ADL. Second, the wording of the MCBS, which asks about receipt of help, not need for help, is dependent on supply and access to help, not only health status. We believe it is inappropriate for a payment model to use a measure of impairment that is confounded by availability of help and the provision of care. We also believe there is a reason to use report of difficulty rather than receipt of help related to the incentives that will follow any risk adjustment methodology. Use of report of difficulty will focus health plans on identifying persons with difficulty and addressing the underlying problems to reduce potential adverse outcomes.

Results. Table 4-6 reports the results of the Functional Status models using both 1991 survey data to predict 1992 costs and 1992 survey data to predict 1993 costs. In both

Table 4-6

Regression of Medicare Payments On Age, Gender, and Functional Status

	1991 Characteristics/ 1992 Payments			1992 Characteristics/ 1993 Payments		
Variable	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic
Intercept	1,978	223	8.86 ***	1,953	219	8.92 ***
Age (65-74 omitted)						
0-64	-1,125	440	-2.55 **	-872	334	-2.61 ***
75-84	447	264	1.70 *	1,003	266	3.77 ***
85+	721	416	1.73 *	1,123	365	3.08 ***
Male	477	237	2.01 **	1,006	224	4.48 ***
Functional Status (no limitations omitted)						
5-6 ADLs	5,589	488	11.45 ***	4,882	446	10.94 ***
3-4 ADLs	3,517	428	8.23 ***	3,648	418	8.74 ***
1-2 ADLs	2,537	294	8.63 ***	2,336	290	8.06 ***
IADLs only	1,129	357	3.17 ***	1,239	324	3.83 ***
R Squared	0.0207			0.0247		
Adjusted R Squared	0.0200			0.0240		
Observations	10,892			10,531		
F	28.71 ***			33.33 ***		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor.

*** Statistically significant at 1 percent level.

** Statistically significant at 5 percent level.

* Statistically significant at 10 percent level.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), and 7(1993).

years, the ADL scale coefficients follow the expected pattern: the more ADL impairments reported, the higher the coefficient. For example, those reporting 5-6 ADL impairments had costs over twice those of persons reporting only 1-2 ADL impairments and 4-5 times the costs associated with IADLs impairments only in both models. The coefficients are also sizable: even the report of only 1-2 ADLs adds over $2,000 to the predicted costs ($2,537 in 1991/1992 and $2,336 in 1992/1993). The most likely tasks in the 1-2 ADL category are mobility or bathing, or a combination of bathing and dressing, mobility and bathing, or mobility and transfers. All of the ADL and IADL categories were statistically significant at the .01 level or better.

Adding disability measures like these functional status variables to the age, gender only model reduces the older age coefficients more than any of the other models presented in this chapter. This indicates that disability, as measured by functional status, is strongly correlated with age. From a theoretical perspective, the reduction in significance of age in the presence of functional status measures indicates that functional status, not age per se, is influencing costs. While gender and age remain statistically significant at least at the .10 level, the coefficients are much lower than for the ADL variables, and less consistently significant at .05 or better levels. Nonetheless, age may be preferable for payment model development because it is readily available, can proxy for functional status on the aggregate level, and is not subject to manipulation.

The R-square for the 1991/1992 model is 0.0207. For the 1992/1993 model, the R-square is 0.0247, only slightly higher, in contrast to the large gain in the self-rated health

status R-square. The functional status coefficients are relatively stable in the two samples, in contrast to some of the other models.

4.2.5 Comprehensive Survey Model

The comprehensive survey model combines the three health status measures used individually in the three previous models: self-rated health status, self-reported chronic conditions, and functional status. It is used to examine the joint performance of these variables, in particular, whether they are redundant or measure independent dimensions of health useful for predicting future medical expenditures.

Although a separate model was not estimated for them, we added physical functioning variables to the comprehensive survey model. Difficulty walking and difficulty lifting were found to have independent predictive power (see the appendix analyzing disability models) and are relatively objective. The MCBS asked if a person "was unable to perform or had/a lot, some, a little, or no" difficulty walking two to three blocks or lifting or carrying a 10 pound object like a sack of potatoes. We coded difficulty walking and lifting as dichotomous variables with 1 indicating "any difficulty" and 0 indicating "no difficulty".

Regression estimates of the comprehensive survey model are reported in Table 4-7. It includes age, gender, self-rated health status, functional status, physical functioning (difficulty walking and lifting), and chronic conditions. Only the conditions selected for the chronic conditions payment model (Table 4-5) are included in the comprehensive survey model. The comprehensive survey model was estimated only on 1991/1992 data.

Table 4-7

Comprehensive Survey Model

	1991 Characteristics/ 1992 Payments		
Variable	Coefficient	Standard Error	t Statistic
Intercept	1,093	331	3.30 ***
Age (65-74 omitted)			
0-64	-1,304	449	-2.90 ***
75-84	433	264	1.64
85+	974	422	2.31 **
Male	392	246	1.59
Self Rated Health Status (excellent omitted)			
Poor	1,890	542	3.49 ***
Fair	579	423	1.37
Good	235	368	0.64
Very Good	121	373	0.32
Functional Status (no limitations omitted)			
5-6 ADLs	2,326	624	3.73 ***
3-4 ADLs	732	526	1.39
1-2 ADLs	764	368	2.08 **
IADLs only	59	376	0.16
Difficulty walking 2-3 blocks	581	320	1.82 *
Difficulty lifting	786	298	2.64 ***
Chronic Conditions			
Arteriosclerosis	551	354	1.55
Heart Attack	1,381	359	3.84 ***
Other Heart Conditions	590	283	2.08 **
High Cost Cancer	1,226	687	1.79 *
Diabetes	1,257	333	3.78 ***
Osteoporosis	787	456	1.73 *
Parkinson's Disease	1,527	925	1.65 *
COPD	908	352	2.58 ***
Partial Paralysis	347	453	0.77
Amputation of arm/leg	3,155	1021	3.09 ***
Lost urine > once per week	680	415	1.64
R Squared	0.0344		
Adjusted R Squared	0.0322		
F-Ratio	15.47 ***		
Observations	10,892		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor.
*** Statistically significant at 1 percent level.
** Statistically significant at 5 percent level.
* Statistically significant at 10 percent level.

SOURCE: Medicare Current Beneficiary Survey, the estimation sample consisting of Rounds 1(1991) and 4(1992).

If the survey health status measures were fully redundant, the R-square from the joint model would be equal to the highest R-square for any individual model. If the measures were orthogonal, the R-squares of the comprehensive model would be the sum of the R-squares of the individual models. Because age and gender are included in all models, we compare R-squares subtracting the R-square (0.37 percent) of the age/gender model (Table 4-1). Since an individual model for difficulty walking and lifting was not estimated, we use the results of the self-rated health, functional status, and chronic conditions models to calculate bounds on the R-squared of the combined model. The 1991/1992 incremental R-squares for the individual models are 2.16 percent (chronic conditions [Table 4-5]), 1.48 percent (self-rated health status), and 1.7 percent (functional status). The sum of the incremental R-squares (including the age/gender model R-square) is 5.71 percent (=2.16% + 1.48% + 1.7% + 0.37%). Hence, if there were no redundancy, the R-square of the comprehensive model would be at least 5.71 percent, but if there were full redundancy it would be 2.53 percent (the R-square of the chronic conditions model).[9] The actual R-square of the comprehensive model is 3.44 percent, closer to the hypothetical "full redundancy" case. In short, although not fully redundant, there is considerable redundancy among the survey health status measures. This implies that there are sharply diminishing returns to asking more health status questions in terms of ability to predict future expenditures.

The redundancy among health status measures is reflected in the coefficients of the comprehensive model. Most are considerably smaller than in the individual health status

[9] "At least" 5.71 percent because difficulty lifting/walking have some explanatory power. The full redundancy bound assumes that the R-square of the unestimated walking/lifting model would be lower.

models as the effect of a dimension of health is now apportioned among several highly intercorrelated variables (compare Table 4-7 to earlier Chapter 4 tables). Several of the coefficients are no longer statistically greater than zero, such as "fair" or "good" self-rated health and partial paralysis. Also, a disabled (under age 65) female reporting excellent health, no activity limitations, no difficulty walking or lifting, and no chronic conditions would be assigned a negative payment. This clearly undesirable situation is symptomatic of highly redundant health status measures, and the need for different parameter estimates for the disabled and elderly.

4.2.6 SF-36-Like Model

We developed a model simulating four of the eight scales from the SF-36 to provide a comparison to our other survey models and to other work done using the SF-36 for risk adjustment (Hornbrook and Goodman, 1995). The SF-36 is a widely-used 36 item health status questionnaire developed to measure outcomes of medical care (Ware, 1993; Ware and Sherbourne, 1992). While the MCBS and SF-36 questions differ in details of wording, we were able to construct simulated scores for the Physical Functioning, General Health, Social Functioning and Role-Physical scales. These are four of the five scales that Hornbrook and Goodman (1995) found to be predictors of medical costs. The details of our simulation of the SF-36 scales are included in an appendix. Our SF-36-like scales have not been tested for equivalence to the actual SF-36 scales.

The SF-36-like model is presented in Table 4-8. Higher scores in the SF-36 are associated with better health. Because higher scores are associated with better health, the SF-36 model has a large intercept and the coefficients are negative. In other words, the coefficients represent reductions in medical payments as health improves. Physical functioning is scored from 0 to 100, with 100 representing the best functioning. As compared to someone with the worst functioning (score = 0), the coefficient of –$34 implies that someone with the best functioning (score = 100) has $3,400 lower Medicare payments, other things equal. General health (which is equivalent to what we have called "self-rated health status") is scored on a 1 to 5 scale, with 5 representing the best health. The coefficient of -$350 implies that someone rating himself or herself in the best general health has $1,400 (=1750 – 350) lower Medicare expenditures than someone in the worst general health, all else equal. Social functioning is scored on a scale of 1 to 4, with 1 indicating social activities limited all of the time, and 4 indicating no limitations. The coefficient of -$658 implies $1,974 (= 2632 – 658) lower expenditures with no limitations compared to limited all of the time, ceteris paribus. Finally, Role Physical is scored on a 2 to 8 scale, with 8 denoting the fewest limitations. Role Physical's coefficient of -$190 means that someone with no limitations has $1,140 (= 1520 – 380) lower expenditures than someone fully limited, holding other regression variables constant. The coefficients of Physical Functioning, General Health, and Social Functioning are all statistically significant at the 1 percent level. Role Physical is significant at the 10 percent level.

Table 4-8

SF-36-Like Model

Variable	1991 Characteristics/ 1992 Payments Coefficient	Standard Error	t Statistic
Intercept	10,876	937	11.61 ***
Age (65-74 omitted)			
0-64	-1,818	435	-4.18 ***
75-84	339	264	1.28
85+	484	428	1.13
Male	692	240	2.88 ***
SF-36 Like Scales[1]			
Physical Functioning (0-100)	-34	6	-5.67 ***
General Health (1-5)	-350	116	-3.02 ***
Social Functioning (1-4)	-658	162	-4.06 ***
Role-Physical (2-8)	-190	113	-1.68 *
R-squared	0.0288		
Adjusted R-squared	0.0281		
F-Ratio	39.98 ***		
Observations	10,845		

[1] Simulated SF-36 scales that have not been tested for equivalence with the actual scales.

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor.

*** Statistically significant at 1 percent level.

** Statistically significant at 5 percent level.

* Statistically significant at 10 percent level.

SOURCE: Medicare Current Beneficiary Survey, the estimation sample consisting of Rounds 1(1991) and 4(1992).

The R-square of the SF-36-like model exceeds that of any of the single-dimension models (self-rated health, functional status, chronic conditions), which is not surprising since the SF-36-like scales combine information from self-rated health and functional status, and add social functioning. The SF-36-like model does not, however, utilize indicators of specific chronic conditions such as heart disease or diabetes, nor does it use the full range of ADLs. Thus, the SF-36 model does not have as great predictive power as the comprehensive survey model discussed in the previous section. Adding chronic condition indicators would be a means of improving the predictive power of the SF-36-like model.

4.2.7 Claims Diagnoses Models

Over the past decade, considerable work has been done to predict Medicare program expenditures using diagnoses recorded on the administrative claims providers submit to the Medicare program. One significant line of research in this area is the Diagnostic Cost Group, or DCG family of models. Recently the DCG models have been updated and extended (Ellis *et al.*, 1996).

Claims-diagnosis-based models are fundamentally different from the survey health status measures because they do not rely on responses Medicare beneficiaries give to questionnaires, but rather the diagnoses providers record on claims. They are related to the MCBS self-reported chronic conditions model (Section 4.2.3 above) in that both models use diagnostic information to predict medical expenditures. However, claims diagnoses are drawn from the comprehensive International Classification of Diseases, Ninth Revision, Clinical

Modification, comprising over 10,000 categories. In contrast, the MCBS queries for about 20 select chronic conditions. A second fundamental difference is that claims diagnoses are typically recorded by professional medical records coders based on physician medical records or instructions, whereas the MCBS diagnoses are reported by beneficiaries.

To compare to the survey health status measures, we constructed predicted 1992 and 1993 expenditure scores from two variants of the DCG models using the Medicare claims linked to the MCBS sample. One DCG variant employed was the DCG/PIP, or PIPDCG, model. It utilizes age, gender, and 12 categories of principal inpatient diagnoses to forecast each individual's expenditures in the next year (Ellis *et al.*, 1996). The other DCG variant employed was the prospective "Hierarchical Coexisting Conditions," or DCG/HCC model (Ellis *et al.*, 1996). This model employs 34 diagnostic categories plus age and gender to predict medical expenditures, using all diagnoses on each individual's hospital and physician claims in the preceding year. These models' parameters (i.e., the predicted expenditure weights associated with their diagnostic categories) were not estimated on the MCBS. That is infeasible given the MCBS's small sample sizes. Instead, we used the estimated coefficients from the 2.5 percent Medicare claims sample (n = 680,000) that Ellis *et al.*, (1996) employed in their model development. The models as parameterized by Ellis *et al.*, 1996, are the ones that are validated in Chapter 5.

For informational purposes, we regressed 1992 and 1993 Medicare expenditures against the PIPDCG and HCC scores.[10] These regressions are reported in Tables 4-9 and 4-9a, respectively. Note that age and gender are not included as additional covariates because they are incorporated into the DCG scores.[11] If the DCG models were perfectly unbiased, the intercepts in Table 4-9 and 4-9a would be zero and the coefficient of the DCG scores would be $3,783, which is the mean expenditure of the sample used to estimate the DCG models (Ellis *et al.*, 1996). Note that this hypothesis cannot be rejected in Table 4-9 and 4-9a: the intercepts are not significantly different from zero (at the 5 percent level), and the DCG score coefficients are not significantly different from $3,783. The R-squares indicate that the claims-diagnosis-based DCG models have greater predictive power than the survey models, and that using all hospital and physician diagnoses (the DCG/HCC model) substantially raises forecasting accuracy as compared to using only principal hospital diagnoses (the PIPDCG model).

---

[10] The "scores" are predicted expenditures divided by mean expenditures in the DCG model development sample of $3,783. Thus, the scores are centered around 1.0.

[11] The DCG models use 12 interacted age/sex cells, a greater number than we included in estimating the survey models. However, little of the DOG models' explanatory power derives from their more detailed demographic cells.

Table 4-9

Regression of Medicare Payments On PIPDCG Score

	1991 Characteristics/ 1992 Payments			1992 Characteristics/ 1993 Payments		
Variable	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic
Intercept	-46	207	-0.22	102	192	0.53
PIPDCG Score	3,889	176	22.10 ***	3,969	165	24.01 ***
R Squared	0.0429			0.0519		
Adjusted R Squared	0.0428			0.0518		
Observations	10,893			10,532		
F	488.34 ***			576.37 ***		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor.

*** Statistically significant at 1 percent level.

** Statistically significant at 5 percent level.

* Statistically significant at 10 percent level.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), and 7(1993).

Table 4-9a

Regression of Medicare Payments On HCC Score

	1991 Characteristics/ 1992 Payments			1992 Characteristics/ 1993 Payments		
Variable	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic
Intercept	-207	177	-1.17	289	167	1.74 *
HCC Score	4,014	138	29.00 ***	3,700	126	29.41 ***
R Squared	0.0717			0.0759		
Adjusted R Squared	0.0716			0.0758		
Observations	10,893			10,523		
F	841.01 ***			864.77 ***		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor.

*** Statistically significant at 1 percent level.

** Statistically significant at 5 percent level.

* Statistically significant at 10 percent level.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), and 7(1993).

4.2.8 Claims Diagnoses Plus Survey Models

A limitation of diagnostic information is that severity of illness is generally not measured.[12] Survey-gathered health status or functional scales may thus usefully augment claims diagnoses by measuring severity of illness. For example, can someone with heart disease walk across the room or not? Alternatively, survey health or disability scales may be largely determined by diagnosis and thus redundant with claims diagnoses. The claims diagnoses plus survey models combine the two sources of information in regression models to determine the incremental gain from adding survey information to claims diagnoses (or vice versa). Only self-rated health and functional status are added to the PIPDCG and HCC scores because the self-reported chronic conditions probably largely duplicate the claims diagnoses already incorporated into the DCG scores.[13] Age and gender are also omitted because they are incorporated in the DCG scores.

Estimates for the combined claims-survey models are reported in Tables 4-10 and 4-10a.[14] Adding the two survey variables to the PIPDCG score raises the R-square from 4.3 percent to 5.2 percent (1991/1992), and from 5.2 percent to 6.7 percent (1992/1993). This

---

[12] ICD-9 codes occasionally do incorporate severity of illness information, and severity may be indirectly measured through comorbidities. Nevertheless, ICD-9 makes no attempt to systematically record severity, and typically severity information is lacking.

[13] Self-reported conditions measure whether someone has "ever been told" they have the condition whereas claims measure treatment of the condition in the last year. Thus, self-report may pick up some conditions that are either untreated or where treatment occurred prior to the last year. In addition, self-reported conditions will pick up diagnoses for non-hospitalized beneficiaries, and comorbidities for hospitalized beneficiaries, not taken account of in the PIPDCG model. Thus, adding self-reported conditions to the PIPDCG model would be a particularly appropriate subject for future research.

[14] Ideally the entire model would be reestimated, including the weights on the DCG diagnostic categories, but that is not possible with MCBS sample sizes.

represents a gain of around one percentage point in R-square, or about 25 percent. Adding self-rated health and functional status to the HCC score raises the R-square from 7.2 percent to 7.5 percent (1991/1992), and from 7.6 percent to 8.2 percent (1992/1993). This is a gain of about one-half percentage point in R-square, or about 7 percent. Not surprisingly, the survey information adds more predictive power to principal inpatient diagnoses (the PIPDCG model) than to all inpatient and ambulatory diagnoses (the DCG/HCC model). The improvement in the PIPDCG model's ability to forecast expenditures is significant, while the improvement in the DCG/.HCC model is modest.

Tables 4-10 and 4-10a indicate that, on average, large groups of people reporting poor health or with multiple activity limitations are substantially more expensive (by thousands of dollars) than predicted by the DCG models.[15] Thus, the survey measures can be used to significantly improve predictions of average expenditure for large groups of beneficiaries reporting poor health or multiple activity limitations. This also implies that providers could use self-rated health or functional status to successfully risk select against the Medicare program even when payments are adjusted by the DCG scores. However, risk selection will not be notably successful except in large beneficiary populations.

In terms of the two survey measures, the results are not consistent across years. In the 1991/1992 sample, functional status appears to be the more significant variable to add to the DCG scores (or to risk select against them), while in the 1992/1993 sample self-rated

[15] A more appropriate way to demonstrate this point is to regress the difference of actual expenditures and expenditures predicted from the HCC model on survey self-rated health and functional limitations. When we did this (not shown), the qualitative results were the same as reported in the text.

Table 4-10

Regression of Medicare Payments On PIPDCG Score, Self-Rated Health Status, and Functional Status

Variable	1991 Characteristics/ 1992 Payments			1992 Characteristics/ 1993 Payments		
	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic
Intercept	-878	325	-2.70 ***	1,186	296	-4.01 ***
Self-Rated Health Status (excellent omitted)						
Poor	2,229	498	4.47 ***	3,864	479	8.07 ***
Fair	1,071	402	2.66 ***	1,926	374	5.15 ***
Good	584	359	1.63	1,552	327	4.75 ***
Very Good	259	368	0.71	755	330	2.29 **
Functional Status (no limitations omitted)						
5-6 ADLs	2,622	505	5.19 ***	2,039	480	4.25 ***
3-4 ADLs	1,384	441	3.13 ***	1,556	439	3.55 ***
1-2 ADLs	1,255	303	4.14 ***	949	289	3.28 ***
IADLs only	235	353	0.67	181	315	0.57
PIPDCG Score	3,369	183	18.45 ***	3,409	170	20.00 ***
R Squared	0.0524			0.0673		
Adjusted R Squared	0.0516			0.0664		
Observations	10,892			10,531		
F	66.81 ***			84.32 ***		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor.

*** Statistically significant at 1 percent level.

** Statistically significant at 5 percent level.

* Statistically significant at 10 percent level.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), and 7(1993).

Table 4-10a

Regression of Medicare Payments On HCC Score, Self-Rated Health Status, and Functional Status

	1991 Characteristics/ 1992 Payments			1992 Characteristics/ 1993 Payments		
Variable	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic
Intercept	-546	301	-1.81 *	-541	284	-1.90 *
Self-Rated Health Status (excellent omitted)						
Poor	1,139	496	2.30 **	2,771	457	6.06 ***
Fair	331	399	0.83	1,119	375	2.98 ***
Good	37	355	0.11	1,009	335	3.01 ***
Very Good	42	363	0.12	521	340	1.53
Functional Status (no limitations omitted)						
5-6 ADLs	1,773	501	3.54 ***	1,074	446	2.41 **
3-4 ADLs	897	437	2.05 **	915	416	2.20 **
1-2 ADLs	961	300	3.20 ***	490	289	1.69 *
IADLs only	73	348	0.21	40	311	0.13
HCC Score	3,688	148	24.86 ***	3,325	137	24.69 ***
R Squared	0.0752			0.0823		
Adjusted R Squared	0.0744			0.0815		
Observations	10,892			10,531		
F	98.37 ***			104.81 ***		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor.

*** Statistically significant at 1 percent level.

** Statistically significant at 5 percent level.

* Statistically significant at 10 percent level.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), and 7(1993).

health status seems to be the more important variable incrementally. This is consistent with the results for these two variables in individual models (Sections 4.2.2 and 4.2.4 above) and is a symptom of the small MCBS sample sizes.

4.2.9 Prior Use Model

It is well established that prior use models perform better (i.e., have higher R squares) than typical age/gender models. Although prior use models often have the best predictive power for future medical care expenditures, they are not regarded as desirable payment models because of several drawbacks. The first drawback, one shared by the DCG/HCC-type models, is that they rely on claims data -- data that are not available for new Medicare beneficiaries -- and require claims-like encounter data from HMOs. A more serious problem is that prior use models systematically over- or under-predict expenditures for given individuals because of regression towards the mean. Regression towards the mean is strongest for persons that suffer from random acute care episodes and is weakest for persons that suffer from chronic conditions. The reason is that medical care expenditures have a transitory component and a permanent component. Welch (1985) and others have associated the transitory component with random acute care episodes and the permanent component with chronic conditions. It is the temporary nature of the transitory component that is responsible for regression towards the mean. A third drawback is that paying more based on higher prior costs may encourage inefficiency, and be unfair to efficient health plans.

Despite concerns about prior use models, a simple version that uses combined Medicare Part A and Part B payments in 1991 to predict 1992 total Medicare payments was estimated for comparison purposes (Table 4-11). Controlling for age and gender, the prior use model predicts that for every dollar of payments in 1991, 45 cents was paid in 1992. The 1992/93 regression has about the same R square as the 1991/92 regression. The regression coefficients, however, are strikingly different. The coefficients for the two oldest age groups become much larger, more than can be simply accounted for by medical care price inflation. The increase in the two age coefficients in 1992/93 might account for the much lower prior use coefficient in the 1992/93 regression than in the 1991/92 regression - 25 percent lower at 33 cents. The 1991/1992 model was used in the validation analysis (Chapter 5).

Table 4-11

Regression of Medicare Payments On Medicare Payments In Previous Year

	1991 Characteristics/ 1992 Payments			1992 Characteristics/ 1993 Payments		
Variable	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic
Intercept	2,190	191	11.47 ***	2,103	196	10.71 ***
Age (65-74 omitted)						
0-64	284	413	0.69	122	309	0.39
75-84	691	255	2.71 ***	1,338	256	5.22 ***
85+	1,847	389	4.74 ***	2,261	336	6.74 ***
Male	38	230	0.17	685	218	3.14 ***
Previous Medicare Payments	0.446	0.017	25.65 ***	0.330	0.014	24.11 ***
R Squared	0.0605			0.0603		
Adjusted R Squared	0.0601			0.0599		
Observations	10,893			10,532		
F	140.29 ***			135.01 ***		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor.

*** Statistically significant at 1 percent level.

** Statistically significant at 5 percent level.

* Statistically significant at 10 percent level.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), and 7(1993).

5

Validation of Risk Adjustment Models

In the previous chapter, we discussed the estimation of various risk adjustment models using MCBS data. In this chapter, we validate the models of 1992 spending predicted by 1991 beneficiary characteristics. Our validation approach is to predict 1993 spending using the 1991/1992 estimated parameters applied to 1992 beneficiary characteristics. We then compare actual 1993 spending with predicted 1993 spending for each of the models developed in the preceding chapter.

By using different years to estimate and validate the risk adjustment models, we can evaluate each model free of bias from "overfitting," which results when regression coefficients are estimated based on small samples in individual cells or when the dependent variable has a skewed distribution. Both of these problems are present for our models since Medicare payment is highly skewed and many of the analytic variables used in estimation have few individuals in each categorical level of the variable. For example, some of the chronic conditions have very few individuals. There are also spending outliers that can affect model estimates. For these reasons, predictive power of a model needs to be judged on an independent validation sample.

Our validation analysis has two significant limitations. The MCBS sample is longitudinal and hence most of the MCBS sample beneficiaries used in estimation are also

included in the validation sample. A nonindependent validation sample should overstate the predictive power of the models when compared with a totally independent validation sample. Unfortunately, the MCBS does not have enough sample size to do a split sample validation analysis. However, due to the low correlation of medical expenditure from one year to the next, a different year of data for the same individuals should provide a reasonable validation. The second limitation is related to the small MCBS validation sample. This leads to random errors in validation results, and the validation results may not generalize to other samples.

5.1 Data and Methods

The validation analysis is performed on the MCBS sample using 1992 (Round 4) and 1993 (Round 7) data. The sample selection criteria were similar to the ones used to obtain the estimation sample in the previous chapter. Briefly, the sample excludes those who died before 1993, HMO enrollees, ESRD patients, those living outside the U.S., and those with missing values for any of the variables used in the analysis. Furthermore, sampled beneficiaries had to be continuously eligible for both Part A and Part B Medicare.

5.1.1 Non-Random Validation Groups and Their Characteristics

Table 5-1 describes the characteristics of the validation sample. The total sample size was 10,532. The table shows the sample sizes for various groups of beneficiaries categorized by demographic, health status, health care utilization and other characteristics in 1992. The

Table 5-1

Sample Characteristics of the 1992 Non-Random Subgroups Used For Model Validation

Validation Group	Sample Size	Percentage[1]	1993 Mean Expenditure[2]
Overall Sample	10,532	100.0 %	$4,097
Age			
0-64	1,658	15.7	3,626
65-74	3,724	35.4	3,165
75-84	3,578	34.0	4,773
85+	1,572	14.9	5,888
Gender			
Female	6,087	57.8	3,949
Male	4,445	42.2	4,294
Medicare Status			
Elderly	8,874	84.3	4,192
Disabled	1,658	15.7	3,626
Institutional Status			
Non-institutionalized	9,763	92.7	3,975
Institutionalized	769	7.3	5,880
Elderly Institutional Status			
Non-institutionalized Elderly	8,226	78.1	4,069
Institutionalized Elderly	648	6.2	6,035
Self Rated Health Status			
poor	1,105	10.5	7,767
fair	2,234	21.2	5,196
good	3,108	29.5	4,186
very good	2,503	23.8	2,884
excellent	1,582	15.0	1,911
Functional Status[3]			
5-6 ADLs	917	8.7	7,705
3-4 ADLs	942	8.9	6,322
1-2 ADLs	2,376	22.6	5,021
IADLs only	1,685	16.0	3,760
None	4,612	43.8	2,735
Chronic Conditions			
Any Chronic Condition	9,671	91.8	4,322
Arteriosclerosis	1,650	15.7	5,776
Heart Attack	1,542	14.6	6,639
Angina	1,608	15.3	6,673
Other Heart Conditions	2,994	28.4	5,958
Hypertension	5,231	49.7	4,556

Table 5-1 (continued)

Sample Characteristics of the 1992 Non-Random Subgroups
Used For Model Validation

Validation Group	Sample Size	Percentage[1]	1993 Mean Expenditure[2]
Chronic Conditions (continued)			
Stroke	1,254	11.9	$5,796
High Cost Cancer	356	3.4	4,835
Low Cost Cancer	1,571	14.9	4,513
Skin Cancer	1,515	14.4	4,392
Diabetes	1,610	15.3	6,310
Rheumatoid Arthritis	1,393	13.2	5,321
Osteoarthritis	5,450	51.7	4,502
Osteoporosis	1,010	9.6	4,633
Mental Retardation	433	4.1	2,252
Dementia	564	5.4	5,399
Mental Disorders	1,021	9.7	4,666
Hip Fracture	610	5.8	5,251
Parkinson's Disease	214	2.0	6,256
Chronic Obstructive Pulmonary Disease	1,459	13.9	5,709
Partial Paralysis	950	9.0	5,364
Amputation of arm/leg	151	1.4	6,791
Lost urine more than once per week	1,669	15.8	6,334
Expenditures, 1993			
Lowest 60 percent	6,322	60.0	277
Fourth Quintile	2,105	20.0	2,517
Fifth Quintile (highest)	2,105	20.0	18,739
Top 5 percent	527	5.0	44,658
Expenditures, 1992			
First Quintile (lowest)	2,108	20.0	1,856
Second Quintile	2,108	20.0	2,422
Third Quintile	2,105	20.0	3,142
Fourth Quintile	2,105	20.0	4,450
Fifth Quintile (highest)	2,106	20.0	9,119
Top 5 percent	527	5.0	13,969
Hospital Admissions, 1993			
no admissions	8,193	77.8	941
one admission	1,528	14.5	10,524
two or more admissions	811	7.7	26,964
Hospital Admissions, 1992			
no admissions	8,645	82.1	3,199
one admission	1,250	11.9	6,594
two or more admissions	637	6.0	12,496

Table 5-1 (continued)

Sample Characteristics of the 1992 Non-Random Subgroups
Used For Model Validation

Validation Group	Sample Size	Percentage[1]	1993 Mean Expenditure[2]
Income			
≤ $15,000	6,131	58.2	$4,427
$15,001 - $25,000	2,198	20.9	3,627
>$25,000	1,979	18.8	3,362
Not Reported	224	2.1	7,100
Supplemental Insurance			
Medicare Only	1,290	12.2	3,416
Medicaid	2,007	19.1	5,473
Other Supplemental Coverage[4]	7,235	68.7	3,849
Education			
< 12 years	4,733	44.9	4,380
= 12 years	3,121	29.6	3,852
> 12 years	2,331	22.1	3,711
Not Reported	347	3.3	5,323
Race			
White	9,145	86.8	4,049
Black	1,142	10.8	4,297
Other	245	2.3	4,950
Living Status			
Living Alone	2,976	28.3	4,368
Living with Spouse	4,755	45.1	3,496
Living with Others	2,801	26.6	4,907

[1] Unweighted percentages of sample observations, i.e., first column divided by 10,532.

[2] Weighted by product of MCBS sampling weight and fraction of the year eligible.

[3] ADL: Activities of Daily Living; IADL: Instrumental Activities of Daily Living.

[4] Other supplemental coverage includes individually purchased (IP), employer sponsored (ES), both IP and ES, and public coverage other than Medicaid, as well as private plans held by a small number of working elderly.

SOURCE: 1992 (Round 4) and 1993 (Round 7) Medicare Current Beneficiary Survey.

table also reports 1993 mean spending for the various subgroups of the validation sample. Average 1993 spending for the total sample was $4,097 in current dollars.[1]

Age, gender, and institutionalization status: Column 1 of Table 5-1 reports the sample sizes and column 2 shows the (unweighted) percentage of beneficiaries in various categories. The highest proportion of the sampled beneficiaries are in the age 65-74 category (35.4 percent). Those under age 65 (the disabled) constitute nearly 16 percent of the sample, and older Medicare beneficiaries above age 85 account for 14 percent of the sample. There are more females in the sample than males (58 versus 42 percent). Only 7 percent of the beneficiaries were in institutions such as nursing facilities.

Average Medicare spending in 1993 for the various subsamples as shown in the last column increases with age from $3,626 for the under-65 to $5,888 for those aged 85 and above. Male beneficiaries accounted for more spending than females on average ($4,294 versus $3,949). The elderly cost about 15 percent more than the disabled on average. Institutionalized persons cost about 48 percent more than the non-institutionalized.

Health status, activity limitations, and chronic conditions: About 10 percent of the beneficiaries reported being in poor health and 15 percent reported excellent health. Others reported intermediate health states (fair, good, or very good). About 56 percent of the sample reported activity limitations with 40 percent reporting limitations in one or more Activities of Daily Living (ADLs). Most people (23 percent), however, had only one or two

[1] Mean spending is annualized, weighted by the inverse of the MCBS sampling probability and adjusted for eligibility duration. The weights are also made to sum to the total sample size.

ADL limitations. More than 90 percent of the sample reported one or more chronic conditions. Hypertension (50 percent) and osteoarthritis (52 percent) were the most common chronic conditions. Other heart conditions and cancers (including skin) were also reported by more than one-quarter of beneficiaries. Between 13 and 16 percent of beneficiaries suffered from arteriosclerosis, heart attack, angina, diabetes, rheumatoid arthritis, pulmonary disease, and urinary difficulties.

As expected, those reporting poor health had larger Medicare spending relative to others with better health. Spending monotonically decreased from $7,767 for those reporting poor health to only $1,911 for those reporting excellent health. Relative to those without any functional limitations, Medicare spending was higher for those with one or more ADLs and more than twice as much for those with 3 or more ADLs. Average Medicare spending varied depending on the chronic condition, ranging from $2,252 for those with mental retardation to $6,791 for those with amputations.

Utilization and expenditure categories: Sampled beneficiaries were categorized based on 1992 and 1993 Medicare expenditures and the incidence and number of hospital admissions. The lowest 60 percent category had an average 1993 spending of only $277. Those in the top 5 percent of spending cost $44,658 on average. The same sample was also categorized based on 1992 spending. High spenders in 1992 also spent higher amounts on average in 1993. Average 1993 spending for the top 5 percent spenders in 1992 was only $13,969 compared with $44,658 for the top 5 percent spenders in 1993, reflecting regression to the mean.

Slightly more than 20 percent of the sample had one or more hospital admissions in 1993. The average Medicare spending for those with one admission was $10,524, and about two and a half times more for those with more than one admission. This compares with an average of only $941 of annual spending for those without a hospital admission. 1993 spending for those who had no admissions in 1992 was much higher ($3,199) than the same year spending for those without admissions in 1993. Those with one or more admissions in 1992, on the other hand, spent lower amounts in 1993 compared with those with admissions in 1993. This reflects the transitory nature of admissions arising largely from acute health events.

Subgroups defined by socio-economic characteristics: Sampled beneficiaries were categorized based on income, education, race, supplemental insurance, and living companionship. More than half of the sampled beneficiaries (58 percent) had an annual income of $15,000 or less.[2] Medicare program spending for them was slightly higher than those with higher income (by about $1,000). Close to 2 percent of the sample did not report their income, and this group had higher Medicare program spending ($7,100) than any other income group. Nearly 45 percent of the sampled beneficiaries did not complete high school. This group cost more than others with at least a high school diploma. Close to 87 percent of the sample was white. Most of the remaining beneficiaries were black (about 11 percent). Program spending was slightly higher for blacks than for whites.

[2] Income appears to be under-reported in the MCBS. See Section 3.5 for discussion.

Medicare supplemental insurance in the form of Medicaid or other coverage (private plans such as Medigap or employer plans, or public coverage other than Medicaid) can impact Medicare spending. Dual Medicare and Medicaid eligibles constituted 19 percent of the validation sample. This group had higher annual spending ($5,473) compared with those not in Medicaid. Those with Medigap or other forms of supplemental insurance spent about 13 percent more than the beneficiaries who did not have any supplemental insurance.

About 28 percent of the sampled beneficiaries were living alone in 1992. Those living with a spouse constituted 45 percent of the sample. This group spent about a $1,000 less than those who lived alone, and about $1,500 less than those living with others including the institutionalized.

5.1.2 Validation Measures for the Various Risk Adjustment Models

1991/1992 risk adjustment regression models as described in Chapter 4 were validated using several statistical measures. The risk adjustment estimation models include the following:

(1) an age-gender model (Table 4-1);

(2) an AAPCC like model (Table 4-2);

(3) a model which includes age, gender, and functional status (Table 4-6);

(4) a model which includes age, gender, and a selected list of self-reported chronic conditions (Table 4-5);

(5) a model with age, gender, and self-rated health status (Table 4-3);

(6) SF-36-like model with four simulated SF-36 scales, age and gender (Table 4-8);

(7) a comprehensive survey model with age, gender, self-rated health status, functional status, and chronic conditions (Table 4-7);

(8) a prior use (expenditures) model (Table 4-11);

(9) two models of DCG scores (PIPDCG and DCG/HCC) based on claims diagnoses (Ellis *et al.*, 1996); and

(10) two combined models of survey and claims variables including the PIPDCG or HCC score, self-rated health status and functional status ADL/IADL scale (Table 4-10).

Risk adjustment models were estimated using 1992 spending and 1991 characteristics. The estimates from the regressions were used to predict 1993 spending using 1992 characteristics of the validation sample. Since the estimation model predicts spending in 1992 dollars, we inflated predicted spending by the ratio of 1993 mean spending to 1992 mean spending. Actual 1993 spending was then compared with predicted spending using three validation measures -- predictive ratios, actual versus predicted spending levels, and R-squares.

Predictive ratios: These are ratios of predicted to actual spending for the subgroup as a whole. Predictive ratios are used to test how well the model predicts spending, on average, for each validation subgroup. A ratio close to 1 indicates better predictive ability for the model than ratios away from 1. For the full sample, the predictive ratio can be different from 1 since the model is being validated on a sample other than the one used for

estimation.[3] Predictive ratios for the non-random validation subgroups were "normalized" by dividing through by the predictive ratio of the full validation sample in order to assess the performance of the model fit for each subgroup relative to the full sample. We also perform hypothesis tests to determine the statistical difference of the predictive ratio from a value of 1.

Comparisons of actual versus predicted mean spending: Mean actual versus predicted spending is compared to assess the predictive accuracy of the models in absolute dollars. This is done for the full sample, and across the non-random subgroups described previously.

R-squares: The R-square measures the proportion of the variance in payments across individuals explained by the risk adjustment model. In that sense, R-squares provide a measure of goodness of fit for model prediction for individuals. We used the following formula to calculate R-squares for each of the various subgroups.

$$R^2 = 1 - \left[\frac{\sum_i w_i \bullet (Actual_i - Predicted_i)^2}{\sum_i w_i \bullet (Actual_i - Mean)^2} \right]$$

where:

i	=	an index over all members of the validation group;
w_i	=	adjusted MCBS weights (see Section 2.3);
$Actual_i$	=	actual payment in 1992 for person i;
$Predicted_i$	=	predicted payment for 1992 for person i from the risk adjustment model; and
Mean	=	weighted mean of actual payment for 1992 for the validation group.

The calculated R-square can be negative if the mean is a better predictor of actual payments than predicted payments.

[3] Note that for the estimation sample, mean of actual and predicted spending will be the same in the full sample, but for other samples, this relationship need not hold.

5.2 Validation Results

5.2.1 Predictive Ratios

Table 5-2 reports predictive ratios for the overall validation sample and validation subgroups. A predictive ratio closer to 1.00 indicates better prediction. A predictive ratio greater than one indicates overprediction, whereas a predictive ratio less than one indicates underprediction. The predictive ratios are subject to random variation because of the limited MCBS sample size. Accordingly, statistical significance of the predictive ratios (difference from 1.00) is indicated in Table 5-2. Among the large number of predictive ratios in Table 5-2, some will be statistically significant by chance. To avoid predictive ratios different than one merely due to random error in predicting overall mean 1993 expenditures, we normalized the predictive ratios by dividing by the predictive ratio for the overall sample.

Predictive ratios are italicized for models that include variables defining the validation group. For example, since functional status is included in the functional status, SF-36-like, comprehensive survey, and combined survey and claims models, the predictive ratios for the functional status validation groups are italicized for these models. While one would expect predictive ratios closer to one for validation groups that are defined by elements of the predictive model, these predictive ratios are still of interest to determine reliability, since estimation and validation are on different years. Moreover, there is no guarantee that models comprising multiple variables will predict well for validation groups defined by a single variable.

Table 5-2

Predictive Ratios for Alternative Risk Adjustment Models, by Validation Subgroup
(A predictive ratio closer to 1.00 indicates better prediction)

Validation Group	Age-Gender	AAPCC[1] Like	Functional Status	Self-Reported Chronic Conditions	Self-Rated Health Status	SF-36 Like	Prior Use	Comprehensive Survey	PIPDCG	PIPDCG & Survey	DCG/HCC	DCG/HCC & Survey
Overall Sample (normalized)[2]	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Age												
0-64	1.04	1.04	0.97	1.01	1.01	0.96	1.08	0.97	1.03	1.28 ***	1.05	1.19 **
65-74	1.07 *	1.07 *	1.05	1.06	1.05	0.95	1.05	1.05	1.03	0.94	1.01	0.96
75-84	0.92	0.92 *	0.94	0.93	0.94	1.02	0.93	0.94	0.98	0.97	0.98	0.98
85+	1.00	0.99	1.03	1.01	1.01	1.05	1.01	1.03	0.96	1.06	0.99	1.05
Gender												
Female	1.04	1.04	1.05	1.05	1.04	1.04	1.05	1.05	1.01	1.03	1.01	1.02
Male	0.94	0.94	0.94	0.93 *	0.95	0.95	0.93 *	0.93	0.99	0.96	0.99	0.97
Medicare Status												
Elderly	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.97	1.00	0.98
Disabled	1.04	1.04	0.97	1.01	1.01	0.96	1.08	0.97	1.03	1.28 ***	1.05	1.19 **
Institutional Status												
Non-institutionalized	1.02	1.02	0.98	1.00	1.01	0.99	1.00	0.99	1.01	0.99	0.99	0.98
Institutionalized	0.77 ***	0.82 **	1.21 ***	1.03	0.86 *	1.15 *	0.96	1.17 **	0.88	1.16 **	1.12	1.27 ***
Self Rated Health Status												
poor	0.49 ***	0.50 ***	0.69 ***	0.68 ***	0.94	0.88 *	0.72 ***	0.93	0.64 ***	0.94	0.76 ***	0.94
fair	0.77 ***	0.78 ***	0.92	0.93	1.00	1.06	0.87 **	1.02	0.85 ***	1.01	0.96	1.03
good	0.97	0.97	0.96	0.96	0.93	0.94	0.95	0.93	0.96	0.94	0.98	0.94
very good	1.42 ***	1.40 ***	1.20 ***	1.22 ***	1.05	1.02	1.23 ***	1.06	1.28 ***	1.03	1.13 **	1.02
excellent	2.15 ***	2.11 ***	1.69 ***	1.64 ***	1.31 ***	1.25 ***	1.77 ***	1.28 ***	1.88 ***	1.29 ***	1.47 ***	1.25 ***
Functional Status[3]												
5-6 ADLs	0.58 ***	0.61 ***	1.10	0.82 ***	0.76 ***	1.02	0.85 **	1.08	0.72 ***	1.06	0.88 *	1.08
3-4 ADLs	0.66 ***	0.67 ***	0.97	0.84 **	0.83 **	1.06	0.83 **	0.94	0.74 ***	0.94	0.85 *	0.95
1-2 ADLs	0.79 ***	0.80 ***	1.03	0.88 **	0.89 **	0.99	0.85 ***	1.03	0.85 ***	1.03	0.90 **	1.03
IADLs only	1.04	1.05	0.95	1.04	1.04	1.04	1.05	0.95	1.06	0.97	1.04	0.96
None	1.44 ***	1.41 ***	0.97	1.21 ***	1.23 ***	0.97	1.22 ***	0.99	1.30 ***	0.98	1.16 ***	0.98
Elderly helped with 3+ ADLs	0.55 ***	0.58 ***	0.71 ***	0.91	0.82 ***	0.94	0.84 ***	0.97	0.70 ***	0.96	0.88 *	1.00
Chronic Conditions												
Any Chronic Condition	0.95 *	0.95 *	0.97	0.99	0.97	0.98	0.98	1.00	0.97	0.99	0.99	1.00
Arteriosclerosis	0.74 ***	0.74 ***	0.87 **	1.08	0.86 **	0.93	0.85 ***	1.09	0.84 ***	0.94	0.95	1.00
Heart Attack	0.61 ***	0.61 ***	0.68 ***	0.96	0.71 ***	0.74 ***	0.77 ***	0.97	0.74 ***	0.79 ***	0.86 **	0.89 *
Angina	0.62 ***	0.63 ***	0.70 ***	0.88 *	0.73 ***	0.76 ***	0.75 ***	0.90 *	0.73 ***	0.80 ***	0.86 **	0.89 *

Table 5-2 (continued)

Predictive Ratios for Alternative Risk Adjustment Models, by Validation Subgroup
(A predictive ratio closer to 1.00 indicates better prediction)

Validation Group	Age-Gender	AAPCC[1] Like	Functional Status	Self-Reported Chronic Conditions	Self-Rated Health Status	SF-36 Like	Prior Use	Comprehensive Survey	PIPDCG	PIPDCG & Survey	DCG/HCC	DCG/HCC & Survey
Chronic Conditions (continued)												
Other Heart Conditions	0.70 ***	0.71 ***	0.78 ***	*0.95*	0.80 ***	0.83 ***	0.83 ***	*0.96*	0.82 ***	0.88 **	0.94	0.96
Hypertension	0.89 ***	0.90 ***	0.93 **	0.97	0.94	0.96	0.93 **	0.99	0.92 **	0.95	0.96	0.97
Stroke	0.73 ***	0.75 ***	0.95	1.01	0.87 **	1.01	0.93	1.06	0.85 ***	1.00	1.01	1.09 *
High Cost Cancer	0.80 *	0.81 *	0.87	*1.18*	0.91	0.95	0.96	*1.20* *	0.97	1.03	1.15	1.17
Low Cost Cancer	0.89 *	0.88 *	0.93	0.96	0.93	0.94	1.00	0.96	0.94	0.96	1.04	1.05
Skin Cancer	0.97	0.95	0.97	1.01	0.98	0.97	0.99	1.00	0.99	0.97	1.02	1.00
Diabetes	0.62 ***	0.64 ***	0.71 ***	*0.95*	0.73 ***	0.77 ***	0.73 ***	*0.96*	0.72 ***	0.80 ***	0.95	0.98
Rheumatoid Arthritis	0.77 ***	0.78 ***	0.91	0.92	0.90	0.98	0.84 ***	1.00	0.81 ***	0.93	0.87 **	0.94
Osteoarthritis	0.93 *	0.94 *	0.99	1.01	0.99	1.03	0.97	1.04	0.95	1.00	0.98	1.00
Osteoporosis	0.95	0.96	1.17 **	*1.39* ***	1.08	1.22 ***	1.14 **	*1.40* ***	1.00	1.17 ***	1.10	1.20 ***
Mental Retardation	1.33 **	1.47 ***	1.39 ***	1.19	1.10	1.21	1.31 **	1.09	1.34 **	1.55 ***	1.35 **	1.47 ***
Dementia	0.94	0.98	1.39 ***	1.21 **	1.12	1.42 ***	1.08	1.38 ***	1.01	1.33 ***	1.21 **	1.38 ***
Mental Disorders	0.81 **	0.84 *	0.89	0.88	0.88	0.93	0.91	0.91	0.86 *	1.02	0.93	1.01
Hip Fracture	0.92	0.94	1.20 **	1.07	1.00	1.20 **	1.12	1.19 **	1.01	1.21 **	1.13	1.24 ***
Parkinson's Disease	0.76	0.77	1.04	*1.32*	0.93	1.08	0.87	*1.34*	0.84	1.05	1.07	1.18
Chronic Obstructive Pulmonary Disease	0.71 ***	0.71 ***	0.80 ***	*1.03*	0.84 ***	0.87 **	0.81 ***	*1.03*	0.83 ***	0.92	0.95	0.99
Partial Paralysis	0.73 ***	0.74 ***	0.98	*1.10*	0.86 *	1.02	0.93	*1.10*	0.81 ***	1.01	0.97	1.08
Amputation of arm/leg	0.57 ***	0.59 ***	0.77	*1.26* *	0.71 **	0.85	0.86	*1.27* *	0.72 **	0.86	0.88	0.96
Lost urine more than once per week	0.70 ***	0.72 ***	0.96	*1.06*	0.83 ***	0.99	0.88 **	*1.07*	0.79 ***	0.98	0.91 *	1.02
Expenditures, 1992												
First Quintile (lowest)	1.99 ***	1.97 ***	1.77 ***	1.64 ***	1.79 ***	1.66 ***	*1.34* ***	1.55 ***	1.56 ***	1.41 ***	0.97	0.92
Second Quintile	1.68 ***	1.66 ***	1.54 ***	1.51 ***	1.55 ***	1.47 ***	*1.16* *	1.44 ***	1.30 ***	1.22 **	1.13	1.10
Third Quintile	1.39 ***	1.39 ***	1.37 ***	1.39 ***	1.36 ***	1.34 ***	*1.01*	1.37 ***	1.09	1.09	1.22 ***	1.22 ***
Fourth Quintile	0.92	0.92	0.95	0.99	0.96	0.97	*0.79* ***	0.99	0.78 ***	0.82 ***	1.02	1.03
Fifth Quintile (highest)	0.46 ***	0.47 ***	0.55 ***	0.56 ***	0.54 ***	0.60 ***	*0.97*	0.61 ***	0.86 ***	0.90 ***	0.89 ***	0.90 ***
Top 5 percent	0.31 ***	0.31 ***	0.42 ***	0.41 ***	0.39 ***	0.47 ***	*1.25* ***	0.47 ***	0.74 ***	0.79 ***	0.86 **	0.88 *
Hospital Admissions, 1992												
no admissions	1.27 ***	1.27 ***	1.23 ***	1.22 ***	1.23 ***	1.20 ***	0.97	1.19 ***	0.99	0.98	1.02	1.02
one admission	0.63 ***	0.63 ***	0.72 ***	0.73 ***	0.70 ***	0.77 ***	1.04	0.78 ***	1.18 ***	1.21 ***	1.04	1.06
two or more admissions	0.33 ***	0.34 ***	0.41 ***	0.44 ***	0.41 ***	0.47 ***	1.07	0.49 ***	0.82 ***	0.85 **	0.86 **	0.87 *
Supplemental Insurance[4]												
Medicaid	0.74 ***	0.91	0.90 *	0.86 ***	0.83 ***	0.93	0.88 **	0.93	0.82 ***	0.98	0.93	1.01
Medicare Only	1.15	1.10	1.17	1.15	1.22 *	1.22 *	1.06	1.20 *	1.13	1.21 *	1.01	1.07
Other Supplemental Coverage	1.05 *	1.01	1.00	1.02	1.02	0.99	1.03	0.99	1.03	0.98	1.02	0.99

Table 5-2 (continued)

Predictive Ratios for Alternative Risk Adjustment Models, by Validation Subgroup
(A predictive ratio closer to 1.00 indicates better prediction)

Validation Group	Age-Gender	AAPCC[1] Like	Functional Status	Self-Reported Chronic Conditions	Self-Rated Health Status	SF-36 Like	Prior Use	Comprehensive Survey	PIPDCG	PIPDCG & Survey	DCG/HCC	DCG/HCC & Survey
Income												
≤ $15,000	0.93 **	0.95	0.96	0.95	0.96	0.98	0.95	0.98	0.95	0.99	0.96	0.98
$15,001 - $25,000	1.09	1.05	1.05	1.06	1.07	1.04	1.04	1.04	1.05	1.00	1.04	1.01
> $25,000	1.19 ***	1.15 **	1.09	1.11 *	1.07	1.03	1.14 **	1.03	1.16 **	1.02	1.11 *	1.04
Not Reported	0.67 *	0.66 *	1.01	0.88	0.79	0.98	0.82	1.00	0.77	1.00	0.92	1.04
Education												
< 12 years	0.92 **	0.94	0.95	0.95	0.98	1.00	0.93 *	0.99	0.94	0.99	0.96	0.99
= 12 years	1.08 *	1.06	1.04	1.05	1.05	1.02	1.06	1.03	1.04	1.00	1.04	1.02
> 12 years	1.09	1.06	1.02	1.02	0.98	0.96	1.08	0.96	1.08	0.98	1.00	0.96
Not Reported	0.86	0.92	1.23 *	1.06	0.96	1.20 *	0.98	1.18	0.95	1.21 *	1.15	1.30 **
Race												
White	1.01	1.00	1.01	1.01	1.00	1.00	1.01	1.00	1.01	1.00	1.01	1.00
Black	0.93	1.00	0.97	0.93	1.01	1.03	0.94	1.00	0.94	1.02	0.97	1.01
Other	0.82	0.88	0.87	0.86	0.89	0.92	0.89	0.92	0.88	0.93	0.89	0.92
Living Status												
Living Alone	0.97	0.97	0.91 *	0.94	0.94	0.90 **	0.95	0.91 *	0.95	0.92	0.95	0.93
Living with Spouse	0.85 ***	0.88 **	0.99	0.93	0.89 **	1.00	0.95	0.99	1.09 **	1.03	0.98	1.05
Living with Others	1.11 ***	1.09 **	1.06	1.08 **	1.10 **	1.07	1.06	1.07	0.91 *	1.03	1.05	1.01

NOTES:
Italics are used to indicate that the validation group is defined by variables included in the predictive model.
1. AAPCC is Adjusted Average Per Capita Cost.
2. Predictive ratios were normalized by dividing them by the predictive ratio of the overall sample.
3. ADL is activity of daily living. IADL is instrumental activity of daily living.
4. Other supplemental coverage includes individually purchased (IP), employer sponsored (ES), both IP and ES, and public coverage other than Medicaid, as well as private plans held by a small number of working elderly.

*** Predictive ratio is significantly different from 1 at the .01 level.
** Predictive ratio is significantly different from 1 at the .05 level.
* Predictive ratio is significantly different from 1 at the .10 level.

SOURCE: 1992 (Round 4) and 1993 (Round 7) Medicare Current Beneficiary Survey.

In the remainder of this section, we highlight findings for selected validation groups of particular interest.

The Institutionalized

The demographic models underpredict spending for the institutionalized, but many of the health status models overpredict spending. This indicates that nursing home residents are absolutely more expensive, but are less expensive to Medicare than community residents with the same diagnoses or functional status. Institutionalized beneficiaries may be less expensive controlling for diagnoses or functional status because of the substitution of nursing home care for the acute care services covered by Medicare.

Self-Rated Health and Functional Status

Not surprisingly, survey models including these variables predict well across validation groups. Other demographic, survey, prior use, and claims-based models are less successful. Comparison of the claims diagnosis DCG models with the combined survey/claims models indicates that survey variables can improve predictions of the claims model across health and functional status groups.

Elderly Receiving Help with ADLs

We included an "elderly receiving help with 3 or more ADLs" validation group since policy makers and providers are interested in the ability of risk adjustment methodologies to

pay accurately for the more functionally impaired elders at risk of institutionalization. Only the combined models, including both the claims-based DCG and the survey measures predict accurately for this group, along with the comprehensive survey model. The self-reported chronic conditions and the SF-36-like model predict reasonably well, even though there are no functional status measures in the first, and limited functional status information in the second. All other models substantially underpredict expenditures for this group.

Prior Utilization

The models using claims information predict payments better for persons with varying levels of prior year payments than the survey or demographic models. The DCG/HCC model underpredicts by only 14 percent among the highest 5 percent prior year spenders as compared to a 53 percent underprediction by the comprehensive survey and SF-36-like models. For the lowest quintile, the DCG/HCC model mispredicts by only 3 percent, versus 64 to 79 percent overprediction by the survey variables. The PIPDCG model does nearly as well as the DCG/HCC model for the highest quintile of prior expenditures, but not as well for the top 5 percent, or for the fourth quintile, and it substantially overpredicts for the bottom 40 percent. The combined survey and claims models do not do much better than the DCG models alone, that is, survey measures do not add much to claims diagnoses for predicting across prior expenditure quintiles. For prior-year hospital admission categories, the prior use model does best, with the two models including the DCG/HCC score a close second, and the two PIPDCG models third.

The Chronically Ill

Across groups of people reporting chronic conditions[4], the models using diagnostic information -- self-reported chronic conditions, comprehensive survey, DCG/HCC, and combined DCG/HCC/survey -- show the fewest statistically significant under- or over-predictions. (The PIPDCG model--which uses only principal inpatient diagnoses--is an exception.) The SF-36-like model also does well, despite utilizing no diagnostic information. The most important chronic condition indicators to include in survey models appear to be heart disease, diabetes, and chronic lung disease. All models overpredict expenditures for the mentally retarded, and all except the demographic models and PIPDCG overpredict for dementia. This could be due to under-provision of care to these groups, or substitution of Medicaid nursing home for Medicare expenditures.

Dual Eligibles

Medicare/Medicaid dual eligibles (identified by "Medicaid" under "Supplemental Insurance" in Table 5-2) are a group of particular interest to state and federal policy makers. Only the combined claims and survey models predicts this group's expenditures accurately. All other models underpredict for this group, although the underpredictions of the AAPCC-like, SF-36-like, comprehensive survey and DCG/HCC models are not statistically significant. Larger sample sizes are needed to confirm these findings.

---

[4] Chronic condition groups could be defined using either survey self-reports or diagnoses recorded on claims. We use self-reports.

Demographic Groups

All the models predict mean expenditures reasonably well across income,[5] education, and race groups, with the exception of the age/gender model. Predicted spending is in general higher than actual spending for beneficiaries who live with individuals other than their spouse, which could reflect substitution of nursing home care for acute medical care, or underservice to these beneficiaries. Living alone, on the other hand, has a tendency to raise actual compared to predicted expenditures.

5.2.2 R-Squares

Overall R-Squares. The last column of Table 5-3 shows the validation R-squares for the various risk-adjustment models for the full validation sample. We also present the estimation adjusted R-squares for the "1992 spending, 1991 characteristics" model which provided the parameter estimates for model validation. These two R-squares are also contrasted with estimation adjusted R-squares based on models which use 1993 spending and 1992 characteristics.[6]

The age-gender and AAPCC like models have validation R-squares of 1 percent or less. Their estimation R-squares are also low relative to other models, suggesting that these models do not explain much of the variation in spending. Among the individual survey-based

[5] See Section 3.5 for a discussion of the limitations of the MCBS income variable.

[6] The comprehensive survey and the SF-36-like models were not estimated on 1992/1993 data. Hence no 1992/1993 estimation R-squares are available for them.

Table 5-3

Percentage of Variation In Medicare Program Payments Explained By Alternative Risk Adjustment Models, By Year and Estimation Versus Validation

	Estimation[a]		Validation
Model	1991/1992	1992/1993	1992/1993
Age & Gender	0.32 %	0.79 %	0.70 %
AAPCC Like	0.45	1.04	0.93
Functional Status	2.00	2.40	2.52
Self-Reported Chronic Conditions	2.40	3.02	2.74
Self-Rated Health Status	1.78	3.16	3.11
Comprehensive Survey	3.22	--	4.18
SF-36-like	2.81	--	4.05
Prior Use (Expenditures)	6.01	5.99	4.13
PIPDCG	4.28	5.18	5.18
PIPDCG and Survey	5.16	6.64	6.56
DCG-HCC	7.16	7.58	7.27
DCG/HCC and Survey	7.44	8.15	7.85

[a] Adjusted R-squares.

NOTE: All models except ones including the DCG scores incorporate age and gender. Age and gender are included within the DCG score. Expenditures are regressed on characteristics from the prior year.

SOURCE: 1991, 1992, 1993 Medicare Current Beneficiary Survey

models (functional status, self-reported chronic conditions, and self-rated health status), the validation R-square is the highest (3.1 percent) for the self-related health status model and lowest (2.5 percent) for the functional status model. The chronic conditions model suffers the largest decline in R-square in validation versus estimation, an indication of overfitting because of small cell sizes for several of the chronic conditions. R-squares are generally higher in the 1992/1993 estimation sample than the 1991/1992 sample, probably because of fewer expenditure outliers in the former sample.

The comprehensive survey model increases estimation and validation R-squares when compared with each of the three individual variable survey-based models (functional status, self-reported chronic conditions, and self-rated health status). However, the incremental increase in explanatory power is not great relative to the individual models. This indicates substantial redundancy among the survey health variables.

Among claims-based models, the validated explanatory power of the DCG models are much higher than the prior-year spending model. The estimation R-squares are more favorable to the prior use model, suggesting overfitting of that model.

The claims-based DCG models have higher explanatory and predictive power than the survey-based models. The DCG/HCC model has about 40 percent greater predictive accuracy than the PIPDCG model. Adding survey measures improves the PIPDCG model significantly, by about 25 percent. Adding survey variables to the DCG/HCC score only modestly increases the validation R-square (from 7.3 percent to 7.9 percent).

R-Squares for Validation Subgroups. Table 5-4 presents validation measures of R-squares for selected validation subgroups. The top row of the table, for the full sample, was discussed in the preceding section. Here we focus on the subgroups. Note that the R-square for a subgroup can be negative if mean spending for that subgroup is a more accurate predictor of individuals' expenditures than predicted expenditures from the model under consideration.

The explanatory power of the claims-based versus survey models differs greatly by aged versus disabled subsamples. For the disabled, the DCG/HCC model is clearly more predictive with an R-square of 14.4 percent versus only 2.6 percent for the comprehensive survey model. Among the elderly, the DCG/HCC model is still better by more than 50 percent, but the gap in R-square is narrowed to 6.7 percent versus 4.3 percent. In addition, survey variables add more predictive power at the margin to claims diagnoses among the elderly. Prior use also does dramatically better among the disabled than the elderly. Expenditures among the disabled are more predictable, and are relatively strongly related to past expenditures and to diagnoses recorded on medical claims, making the disabled particularly suitable for claims-based risk adjustment.

Consistent with its greater overall predictive power, the claims-based DCG/HCC model predicts better among individuals in most subgroups than the survey or prior use models. The DCG/HCC model tends to do better at predicting expenditure differences

Table 5-4

Explanatory Power (R-Square) of Alternative Risk Adjustment Models, by Selected Validation Subgroup

Validation Group	Age-Gender	AAPCC[1] Like	Functional Status	Self-Reported Chronic	Self-Rated Health	SF-36 Like	Prior Use	Comprehensive Survey	PIPDCG	PIPDCG & Survey	DCG/HCC	DCG/HCC & Survey
Overall Sample	0.70%	0.93%	2.52%	74[illegible]	3.11%	4.05%	4.13%	4.18%	5.18%	6.56%	7.27%	7.85%
Age												
0-64	0.0	0.6	0.1	2.1	0.1	1.6	18.5	2.6	9.8	9.4	14.4	13.8
65-74	0.0	0.5	2.6	3.0	3.4	4.4	1.6	5.0	5.1	7.1	8.6	9.3
75-84	0.0	-0.1	1.8	1.7	2.1	3.0	2.9	2.9	3.5	4.8	4.1	4.8
85+	0.0	0.3	0.5	0.4	1.9	2.6	1.8	1.6	3.2	4.3	5.1	5.6
Gender												
Female	0.8	1.1	3.1	3.1	3.6	4.8	5.4	4.8	6.9	8.3	8.2	8.8
Male	0.5	0.7	1.7	2.3	2.5	3.2	2.6	3.5	3.2	4.5	6.1	6.7
Medicare Status												
Elderly	0.7	0.9	2.7	2.8	3.3	4.2	3.0	4.3	4.8	6.3	6.7	7.4
Disabled	0.0	0.6	0.1	2.1	0.1	1.6	18.5	2.6	9.8	9.4	14.4	13.8
Institutional Status												
Non-institutionalized	0.7	0.9	2.5	2.7	3.1	3.9	4.3	4.2	5.1	6.5	7.1	7.8
Institutionalized	-2.9	-2.6	-1.2	-1.6	-0.9	2.3	-2.7	-0.1	2.9	4.0	5.6	5.2
Self Rated Health Status												
poor	-5.6	-5.1	-0.8	-1.4	1.1	2.7	5.6	2.0	4.6	7.8	6.7	8.1
fair	-0.7	-0.6	-0.3	1.3	0.5	1.1	1.8	1.7	4.2	4.6	6.8	6.8
good	0.6	0.6	0.9	1.9	0.4	1.7	0.8	2.3	2.9	3.2	5.7	5.9
very good	-0.6	-0.4	0.5	-0.7	0.8	0.9	0.9	0.7	0.8	1.9	0.6	1.3
excellent	-10.3	-9.4	-1.4	-4.0	1.0	2.4	-5.6	1.9	-4.6	2.2	-0.1	2.5
Functional Status[2]												
5-6 ADLs	-7.5	-6.3	0.0	0.2	-0.3	3.5	1.6	3.1	4.5	9.0	7.3	9.0
3-4 ADLs	-2.0	-2.0	0.6	1.2	1.3	2.9	4.2	2.7	3.0	4.8	7.2	7.8
1-2 ADLs	-0.8	-0.6	0.3	0.3	0.5	1.8	3.3	1.7	2.8	3.9	4.9	5.4
IADLs only	-0.3	-0.2	0.1	0.4	1.0	1.8	0.1	1.9	5.5	5.8	6.3	6.5
None	-1.0	-0.8	0.3	0.6	1.3	1.2	1.0	1.7	1.7	3.0	3.5	4.1
Elderly helped with 3+ ADLs	-7.1	-6.5	-0.8	-2.2	-0.9	2.2	0.9	1.0	1.0	4.6	5.8	7.0
Expenditures, 1992												
First Quintile (lowest)	-4.5	-4.6	-1.5	-3.4	-1.9	-1.2	-0.4	-1.2	-1.1	0.4	-0.1	0.6
Second Quintile	-2.4	-2.5	-1.9	-2.0	-1.7	-1.8	-0.1	-1.4	-0.3	-0.3	0.0	-0.2
Third Quintile	-0.7	-0.4	-1.2	-5.9	-0.5	-0.2	1.2	-3.8	1.1	1.4	-1.2	-0.7
Fourth Quintile	0.2	0.7	1.0	0.6	1.6	1.8	0.0	1.8	-2.1	-0.3	3.1	3.6
Fifth Quintile (highest)	-9.3	-8.9	-5.9	-3.1	-4.7	-2.6	-3.0	-1.4	1.0	3.4	4.5	5.5
Top 5 percent	-21.3	-20.7	-14.9	-13.0	-14.0	-9.9	-15.7	-9.3	0.2	3.2	2.9	4.4
Hospital Admissions, 1992												
no admissions	0.2	-0.1	1.5	1.2	0.2	2.0	2.8	1.6	0.8	2.2	3.3	3.8
one admission	-3.4	-3.5	-1.0	-1.7	-0.4	0.0	-2.9	0.9	-0.2	1.1	3.5	4.1
two or more admissions	-17.8	-18.1	-12.0	-13.9	-9.9	-9.1	-9.1	-7.7	2.4	4.0	1.9	2.8

NOTE: R-square is negative if mean payment is a better predictor than predicted payments for a subgroup.

1. AAPCC is Adjusted Average Per Capita Cost.

2. ADL is activity of daily living. IADL is instrumental activity of daily living.

SOURCE: 1992 (Round 4) and 1993 (Round 7) Medicare Current Beneficiary Survey.

Risk\Reports\tables\chap5\tab5-4.wpd

among individuals in poorer health than among those in better health. This is consistent with its emphasis on multiple, serious, high-cost conditions (Ellis *et al.*, 1996). Adding survey measures (in the combined survey/claims model) improves the ability of the DCG/HCC model to predict expenditure differences among individuals in relatively good health, as well as differences among individuals in the worst health. Quite often, the models are less successful at predicting expenditure differences among individuals in a subgroup than among the overall sample. This is because much of the models' overall explanatory power results from predicting differences among groups.

5.3 Conclusions

In this chapter, we presented and discussed the validation properties of the risk adjustment models defined in Chapter 4. We used predictive ratios to examine how well the risk adjustment models predict on average for nonrandom groups, and R-square measures to examine how well they predict for individuals. Both survey and claims based risk adjustment models performed better than pure demographic models (age/gender and AAPCC like models). The DCG claims-based models performed better than the survey models, both in terms of R-square and across most nonrandom groups. The claims-based models --DCG and prior expenditures -- were particularly better at predicting spending across prior expenditure and utilization groups, and for the disabled. As more variables are added to a model, it becomes more powerful, but at a sharply diminishing rate. Thus, the comprehensive survey model was somewhat, but not greatly, better than any of the individual survey models, and

the combined claims/survey models were only modestly better than the claims-diagnosis models alone.

Incorporating any of the survey or claims-based variables into Medicare payment will reduce risk selection possibilities as compared to the AAPCC. But any of the models we examined is still subject to significant risk selection. Health plans could use the other variables analyzed to profit at Medicare's expense even if Medicare payment were adjusted by the most powerful individual model, the claims-diagnosis-based DCG/HCC model. Hence, risk adjustment should improve risk selection problems, but not eliminate them. Further improvement of risk adjusters is necessary, as well as other policies to prevent risk selection, such as open enrollment.

6 Preferred Survey Models Estimated on Multiple MCBS Years

The analysis in Chapter 4 showed considerable instability in estimates of risk adjustment models across years of Medicare Current Beneficiary Survey (MCBS) data. Adding additional years of data are necessary to obtain more stable estimates of model parameters. In this chapter we use three paired years of MCBS data to obtain more stable estimates.

Rather than reestimating all the models formulated in Chapter 4, we focus on developing a single or preferred survey risk adjustment model. The Health Care Financing Administration (HCFA) is required to risk adjust payments to health plans by January 1, 2000 to account for differences in the health status of enrollees in different plans and in fee-for-service versus capitated plans. The Health of Seniors (HoS) survey provides one vehicle for risk adjusting payments. HoS is being administered to Medicare beneficiaries enrolled in capitated health care plans, providing data on enrollees' health status and other characteristics. However, initially, at least, risk adjustment models will have to be calibrated using another data source, namely the MCBS. The goal of this chapter is to obtain stable parameter estimates using the MCBS for a survey risk adjustment model that can be implemented from HoS. To further this goal, where possible, we have formulated the MCBS

variables we utilize for our preferred survey model to be consistent with questions included on the HoS survey at the time this research was conducted.[1]

6.1 Data and Methods

We utilize four years of MCBS data: Round 1 (1991), Round 4 (1992), Round 7 (1993), and Round 10 (1994). Health status information from the prior year is used to predict total annualized Medicare payments in a year: 1991 health status is used to predict 1992 payments, 1992 health status is used to predict 1993 payments, and 1993 health status is used to predict 1994 payments. Thus, we have three samples of paired years, with roughly 10,000 observations per sample. Payments for 1993 and 1994 were deflated by the ratio of mean payments in each of those years to mean 1992 payments to put all payments in consistent 1992 dollars.

All of our analyses presented here are for pooled years, with a sample size of 32,335. We analyze both the elderly (sample size = 27,130) and the disabled (sample size = 5,205). The MCBS is a panel (longitudinal) survey, with individuals resurveyed in each round. Because medical expenditures for individuals are (positively) correlated over time, our effective sample size is less than our apparent sample size. To account for the panel nature of the MCBS, we present estimates of our final models using a statistical procedure that corrects for the intertemporal intraperson correlation. All models were estimated using the MCBS sampling weights, but they are not corrected for complex sample design features. All

[1] The HoS survey instrument was not finalized at the time this research was conducted.

of our analyses are presented on our single estimation sample. Unfortunately, we do not have available to us an independent validation sample to test for overfitting.

A few individuals had extremely large annualized payments. One individual in 1992 had payments of more than \$2 million, and one had payments of about \$800,000. In 1994, one individual had annualized payments of about \$1 million. No other observations had payments of more than \$500,000 in any year. We estimated models deleting these three observations to test the sensitivity of our results to "outliers" on our dependent variable. With the "outliers" dropped, R-squares rose and coefficient standard errors fell, but parameter estimates were little changed. We conclude that for payment purposes our models are not greatly affected by "outliers" since predicted payments depend on parameter estimates, not on R-squares or standard errors. Since occasional extremely high payments are an inherent part of the medical expenditure distribution, we included the "outliers" in estimating all models presented in this memorandum.

6.2 Variable Selection, Construction, and MCBS/HoS Crosswalk

The MCBS collects a large number of variables that could be used for risk adjustment. Based on our analyses in Chapters 4, 5, and the appendix on functional status models, on comparability with HoS items, and on discussions with HCFA staff, we chose a limited set of variables for this analysis. The criteria for choosing variables were statistical power in predicting future medical care expenditures, availability on the HoS, and suitability for use in

payment. The latter criterion includes considerations such as objectivity, verifiability, widespread acceptance, stability, reliability, and simplicity.

The following survey variables were chosen:

- self-rated (general) health status;
- functional status, measured by Activities of Daily Living (ADLs);
- physical functioning, measured by difficulty in walking and lifting; and
- self-reported chronic conditions.

A crosswalk of these items on the MCBS and HoS is included in Table 6-1. General health was measured on the ordinal scale "excellent, very good, good, fair, poor". Unlike the SF-36, the MCBS requests general health "compared to others your own age". A November 24, 1997 draft of the HoS included both versions of the general health question. Separate dummy variables for each level of general health were created, with the "excellent" category omitted in regression modelling.

Functional status was measured by the ordinal count scale "5-6 ADLs, 3-4 ADLs, 1-2 ADLs, no ADLs". ADLs were measured as dichotomous variables, with respondents reporting difficulty doing because of a health or physical problem considered "limited". The November 24, 1997 draft of the HoS includes the same six ADLs as the MCBS (bathing, dressing, eating, walking, getting in and out of chairs, using the toilet). Each is measured on a three point scale: unable to do, have difficulty doing, no difficulty. By collapsing the HoS "unable to do" and "have difficulty" categories, the ADL scale we used can be constructed from the HoS responses.

Table 6-1 (continued)

Crosswalk Between MCBS and HoS Variables

MCBS (Round 1)	HoS (11/24/97 draft)
Chronic Conditions	Chronic Conditions
Please tell me if a doctor ever told you that you had any of these conditions:	Has a doctor ever told you that you had:
- myocardial infarction or heart attack?	- a myocardial infarction or heart attack?
- angina pectoris or coronary heart disease?	- angina pectoris or coronary artery disease?
- other heart conditions such as congestive heart failure, problems with the valves in the heart, or problems with the rhythm of your heartbeat?	1) congestive heart failure? 2) other heart conditions, such as problems with heart valves or the rhythm of your heartbeat?
- any other kind of cancer (except skin), malignancy or tumor (include benign or nonmalignant tumors or growths)?	- any cancer (other than skin cancer)?
- diabetes, high blood sugar, or sugar in the urine?	- diabetes, high blood sugar, or sugar in the urine?
- emphysema, asthma or COPD?	- emphysema, asthma, or COPD?

The MCBS offers five levels of response to the physical functioning variables "difficulty walking or lifting?": unable to do/a lot/some/a little/no difficulty. On HoS, respondents can answer "walk several blocks?" (equivalent to the MCBS walking question) and "lift or carry groceries?" (equivalent to the MCBS lifting question) with three responses: limited a lot, limited a little, and not limited. For comparability with HoS, we collapsed the five MBS response categories to a tri-level scale "unable to do/a lot of difficulty, some/a little difficulty, no difficulty". Separate dummy variables for "unable to do/a lot of difficulty" and "some/a little difficulty" were constructed for both lifting and walking, with "no difficulty" omitted.

Chronic conditions are measured on the MCBS by asking respondents "has a doctor ever told you that you had [condition]". If a respondent reported a condition in one round, they were coded with the condition in all subsequent rounds. The HoS uses the same question for chronic conditions, although it employs a somewhat shorter and different list of conditions. As discussed in the next section, we focus on the following six chronic conditions in our final model: heart attack, angina, other heart conditions, cancer (except skin), diabetes, and chronic obstructive pulmonary disease (COPD). All of these conditions are available from the HoS. The MCBS "other heart conditions" comprises two HoS questions, for congestive heart failure and for "other" heart conditions. Unlike the HoS, the MCBS cancer question instructs the interviewer to include benign tumors. A separate dummy variable was created for each chronic condition.

Although it has some incremental predictive power, and is available on both the MCBS and HoS, we did not include "social functioning" in our final model. This was because of concerns about its lack of specificity and objectivity for use in a payment model. Nor did we include any Instrumental ADLs, because the HoS does not include them, and because of concerns about their suitability for payment. For example, one of the IADLs is meal preparation, which many men answer that they don't do for nonhealth reasons. Also, not all IADLs were asked for MCBS facility residents. Thus, the functional status scale used in our final model differs from the scale analyzed in Chapter 4 and 5 because it excludes the "IADL only" category.

6.3 Model Estimation and Selection

6.3.1 Exploratory Model with All Chronic Conditions, Aged and Disabled

Table 6-2 presents estimates of alternative survey health status models using our sample of three paired years of the MCBS. We began by estimating (using weighted least squares) an exploratory model for combined aged and disabled with the complete set of MCBS chronic conditions. This is Model (1) in Table 6-2 ("Aged and Disabled, All Chronic Conditions"). Our goal was to determine which set of chronic conditions has predictive power in addition to the other risk adjustment variables. The chronic conditions which have positive and statistically significant coefficients are: heart attack, angina, other heart conditions, cancer, diabetes, mental disorders, chronic obstructive pulmonary disease (COPD), amputation of arm/leg, and incontinence (lost urine more than once per week or on

Table 6-2

Alternative Survey Risk Adjustment Models

Variable	(1) Aged and Disabled All Chronic Conditions[1]			(2) Aged and Disabled Selected Chronic Conditions[1]			(3) Aged Only Selected Chronic Conditions Weighted Least Square[1]			(4) Aged Only Selected Chronic Conditions Variance Components[2]			(5) Disabled Only Selected Chronic Conditions[1]		
	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic
Intercept	813	188	4.33 ***	757	176	4.30 ***	704	188	3.74 ***	790	204	3.87 ***	527	577	0.91
Age (65-74 omitted)															
0-64	-1,515	258	-5.86 ***	-1,341	229	-5.85 ***	-	-	-	-	-	-	-	-	-
75-84	653	142	4.59 ***	628	142	4.43 ***	609	149	4.08 ***	661	166	3.99 ***	-	-	-
85+	1,072	231	4.64 ***	934	226	4.12 ***	901	240	3.75 ***	1,134	268	4.23 ***	-	-	-
Male	583	134	4.36 ***	614	129	4.74 ***	662	143	4.62 ***	658	164	4.02 ***	222	287	0.77
Self-Rated Health Status (excellent omitted)															
Poor	2,549	304	8.39 ***	2,605	301	8.65 ***	3,049	350	8.70 ***	2,686	360	7.46 ***	1,588	648	2.45 **
Fair	1,007	230	4.38 ***	1,009	228	4.43 ***	965	250	3.85 ***	958	259	3.70 ***	1,107	619	1.79 *
Good	614	196	3.13 ***	598	195	3.07 ***	549	210	2.61 ***	547	218	2.51 **	921	611	1.51
Very good	284	196	1.45	272	196	1.39	254	210	1.21	311	214	1.45	645	675	0.96
Functional Status (no ADLs omitted)															
5-6 ADLs	2,152	355	6.07 ***	2,066	324	6.37 ***	1,942	369	5.26 ***	1,687	386	4.37 ***	2,453	584	4.20 ***
3-4 ADLs	923	286	3.22 ***	936	281	3.33 ***	1,016	323	3.15 ***	860	328	2.62 ***	385	479	0.80
1-2 ADLs	486	187	2.59 ***	466	186	2.50 **	511	208	2.45 **	510	210	2.43 **	-21	371	-0.06
Difficulty walking 2-3 blocks ("no" omitted)															
Unable/ a lot of difficulty	1,426	228	6.25 ***	1,463	227	6.45 ***	1,583	255	6.22 ***	1,567	261	6.01 ***	359	452	0.79
Some/ a little difficulty	420	180	2.33 **	415	180	2.31 **	504	198	2.54 **	471	201	2.34 **	-504	410	-1.23
Difficulty lifting ("no" omitted)															
Unable/ a lot of difficulty	952	218	4.37 ***	950	217	4.38 ***	940	244	3.84 ***	865	249	3.47 ***	970	419	2.31 **
Some/ a little difficulty	472	171	2.75 ***	444	171	2.60 ***	439	189	2.33 **	392	190	2.06 **	519	377	1.38
Chronic Conditions															
Arteriosclerosis	39	190	0.20	-	-	-	-	-	-	-	-	-	-	-	-
Heart Attack	1,212	200	6.06 ***	1,249	198	6.31 ***	1,297	217	5.98 ***	1,169	244	4.79 ***	389	483	0.80
Angina	631	201	3.13 ***	635	[illegible]	3.23 ***	553	214	2.58 ***	657	240	2.74 ***	1,782	493	3.62 ***
Other Heart Conditions	623	151	4.12 ***	618	150	4.12 ***	534	164	3.55 ***	571	183	3.11 ***	955	366	2.61 ***
Hypertension	17	128	0.13	-	-	-	-	-	-	-	-	-	-	-	-
Stroke	257	226	1.14	-	-	-	-	-	-	-	-	-	-	-	-
Skin Cancer	29	175	0.17	-	-	-	-	-	-	-	-	-	-	-	-
Cancer, except skin	424	165	2.57 **	462	164	2.82 ***	499	179	2.79 ***	457	201	2.27 **	100	410	0.24
Diabetes	1,403	176	7.99 ***	1,457	174	8.37 ***	1,396	192	7.27 ***	1,476	217	6.80 ***	1,938	388	5.00 ***
Rheumatoid Arthritis	-81	192	-0.42	-	-	-	-	-	-	-	-	-	-	-	-
Osteoarthritis	-237	129	-1.84 *	-	-	-	-	-	-	-	-	-	-	-	-
Osteoporosis	-142	228	-0.62	-	-	-	-	-	-	-	-	-	-	-	-
Mental Retardation	-114	450	-0.25	-	-	-	-	-	-	-	-	-	-	-	-
Mental Disorders	587	268	2.19 **	-	-	-	-	-	-	-	-	-	-	-	-
Alzheimer's	-1,759	362	-4.86 ***	-	-	-	-	-	-	-	-	-	-	-	-
Hip Fracture	-392	303	-1.29	-	-	-	-	-	-	-	-	-	-	-	-
Parkinson's Disease	451	477	0.95	-	-	-	-	-	-	-	-	-	-	-	-
COPD	809	184	4.40 ***	805	183	4.39 ***	847	205	4.12 ***	935	232	4.04 ***	440	362	1.21
Partial Paralysis	-138	260	-0.53	-	-	-	-	-	-	-	-	-	-	-	-
Amputation of arm/leg	1,859	559	3.32 ***	-	-	-	-	-	-	-	-	-	-	-	-
Lost urine > once per week	647	208	3.12 ***	-	-	-	-	-	-	-	-	-	-	-	-
R Squared	0.0472			0.0456			0.0469			-			0.0424		
Adjusted R Squared	0.0461			0.0450			0.0461			-			0.0385		
Observations	32,335			32,335			27,130			27,130			5,205		
F, Chi-Squared	44.4 ***			73.6 ***			66.7 ***			359.2 ***			12.8 ***		
Computer Output	RSK4800.OUT (last model)			RSK4808.OUT (1st Model)			RSK4808.OUT (3rd Model)			RSK4810.OUT (Run #1)			RSK4808.OUT (Run #2)		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor
*** Statistically significant at 1 percent level, ** statistically significant at 5 percent level, and * statistically significant at 10 percent level
[1] Weighted least squares
[2] A variance components model was estimated using SAS PROC Mixed, which does not produce R-squares

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), 7(1993) and 10(1994)

Table 6-2 (continued)

Alternative Survey Risk Adjustment Models

Variable	(6) Disabled Only Collapsed Health Status Variables Weighted Least Squares[1]			(7) Disabled Only Collapsed Health Status Variables Variance Components[2]		
	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic
Intercept	1,351	272	4.97 ***	1,782	344	5.18 ***
Male	280	281	1.00	119	363	0.33
Self-Rated Health Status (all other omitted)						
Poor/Fair	800	335	2.39 **	575	338	1.70 *
Functional Status (all other omitted)						
5-6 ADLs	2,229	496	4.50 ***	1,446	524	2.76 ***
Difficulty walking 2-3 blocks (all other omitted)						
Unable/ a lot of difficulty	854	336	2.54 **	762	348	2.19 **
Difficulty lifting (all other omitted)						
Unable/ a lot of difficulty	801	346	2.32 **	652	341	1.91 *
Chronic Conditions						
Heart Disease	1,425	311	4.58 ***	1,363	379	3.59 ***
Diabetes	2,152	385	5.58 ***	2,122	472	4.49 ***
COPD	632	361	1.75 *	563	439	1.28
R Squared	0.0359			-		
Adjusted R Squared	0.0344			-		
Observations	5,205			5,205		
F, Chi-square	24.2 ***			401.4 ***		
Computer Output	RSK4812.OUT			RSK4813.OUT (1st Model)		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor

*** Statistically significant at 1 percent level.
** Statistically significant at 5 percent level.
* Statistically significant at 10 percent level

[1] Weighted least squares.
[2] A variance components model was estimated using SAS PROC Mixed, which does not produce R-squares

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), 7(1993) and 10(1994).

urinary catheter). We eliminated all other chronic conditions from further models.[2] This does not imply that the eliminated conditions have no relationship to health care costs, only that with the sample sizes available to us from the MCBS, we cannot obtain stable and precise estimates of this relationship.

We also eliminated several of the positive, statistically significant chronic conditions from further models. "Mental disorders", in our opinion, is too broad and vaguely worded on the MCBS to be used for payment purposes. It could range from schizophrenia to mild anxiety. It is only marginal statistically significant in the combined sample, and was not statistically significant in each of the three years (not shown). In addition, when we decomposed the sample by elderly versus disabled (not shown), mental disorders was significant in the elderly equation, but not in the disabled equation, which is surprising and inconsistent with what is found in claims data (Pope et al., 1998). Amputation is reported by only about one percent of the sample. Thus, even with the combined years sample, it is subject to instability from small sample sizes, and is not as useful as other more widespread diagnoses such as heart disease and diabetes. Incontinence is more a symptom than an underlying medical condition, and is difficult to verify.

Other than the chronic conditions, all other health status variables have plausible, statistically significant coefficients (except for "very good" self-rated health). Within each ordinal scale, levels indicating poorer health or greater difficulty are associated with larger increases in future Medicare payments, as expected. For example, a person reporting "poor"

[2] It was surprising to us that the condition "stroke" was not statistically significant. We believe this may be because expensive stroke cases have significant disability, and thus are captured by the health status and disability measures.

health status has, on average, $2,549 of increased Medicare payments the next year compared to $923 for a person reporting "fair" health status. The model adjusted R-square is 4.6 percent, which is good compared to the range of survey models we considered in Chapters 4 and 5.

6.3.2 Model with Selected Chronic Conditions, Aged and Disabled

The second model we present includes only selected chronic conditions (Model (2), "Aged and Disabled, Selected Chronic Conditions" in Table 6-2). It is estimated for the combined aged and disabled sample using weighted least squares. Eliminating all but six of the chronic conditions lowers the adjusted R-square only from 4.61 percent to 4.50 percent, indicating that a simplified chronic conditions list sacrifices little explanatory power. The included chronic conditions of heart attack, angina, other heart conditions, cancer, diabetes, and COPD all have large, positive, and statistically significant coefficients.[3] These conditions are relatively common in the Medicare population and have good clinical face validity as significant, underlying diagnoses affecting long-term health status. They are very likely to be associated with elevated Medicare expenditures in other samples, and their marginal effects on Medicare payments can be estimated reasonably precisely.

Limiting the list of chronic conditions has little effect on other coefficients. The health status coefficients continue to be plausible and significant. However, the coefficient for the

[3] The coefficient on "cancer" is lower than might be expected. This is probably because the MCBS instructs the respondent to include benign neoplasms in addition to malignant neoplasms, and because the cancer may have occurred in the past and now be cured or in remission.

disabled subpopulation (age 0-64) continues to be highly negative. It implies a negative payment for a disabled female reporting excellent health, no limitation in ADLs, no difficulty walking or lifting, and no chronic conditions ($757 -$1,341 = -$584). Although such a person in apparently good health may be rare in the Medicare disabled population, they are not impossible, and it is not desirable for a payment model to predict negative payments for anyone.

More basically, the highly negative "disabled" intercept indicates that the model is overpredicting payments for the disabled on average. This is presumably because certain health or disability states are associated with smaller future Medicare payments among the disabled than among the elderly. Indeed, some evidence of this can be seen by comparing coefficients in Model (3) "Aged Only, Selected Chronic Conditions" to Model (5) "Disabled Only, Selected Chronic Conditions". In the combined model, the elderly (who comprise 90% of the sample) dominate, leading to overprediction for the disabled. Different underlying marginal effects for certain variables would suggest interaction terms on these variables for the disabled. Unfortunately, sample sizes for the disabled are not sufficient to clearly identify a subset of variables for which elderly and disabled coefficients differ. Thus, we decided to estimate further models separately for the disabled and elderly.

6.3.3 Model with Selected Chronic Conditions, Aged Only

Model (3) comprises the same set of variables as Model (2), but it is estimated on the elderly sample only (excluding disabled). Not surprisingly, since the elderly dominate the

combined sample, Model (3) coefficients are similar to Model (2) coefficients. All coefficients in this model are plausible and have face validity.

6.3.4 Model with Selected Chronic Conditions, Aged Only, Estimated by Variance Components

Model (4) is identical to Model (3), except that Model (4) is estimated using a variance components technique ("random effects") in the SAS statistical procedure "PROC MIXED". This technique adjusts for the correlation among observations introduced because the same people are resurveyed over time in the MCBS. As expected, the corrected estimates assign larger standard errors, and thus less statistical significance to the variables because the correction reduces the effective sample size. However, the increase in standard errors is very limited. Coefficients are also not greatly affected. There seems to be a tendency for the larger coefficients to be reduced, perhaps because the sickest and thus most costly individuals have the largest correlation of expenditures over time.

Model (4) is our final, preferred survey health status model for the elderly. It includes an appropriate and defensible set of variables; its coefficient estimates have face validity and we believe will be reasonably stable in other samples; it has good ability to predict future health care costs; it is parameterized using an appropriate statistical technique; and it can be implemented using responses to the HoS survey. Model (4) is repeated in Table 6-3 for better readability.

Table 6-3

Preferred Survey Risk Adjustment Model for Elderly

Dependent Variable: Annual Medicare Payments Per Beneficiary, 1992 Dollars

Variable	Coefficient	Standard Error	t Statistic
Intercept	790	204	3.87 ***
Age (65-74 omitted)			
75-84	661	166	3.99 ***
85+	1,134	268	4.23 ***
Male	658	164	4.02 ***
Self-Rated Health Status (excellent omitted)			
Poor	2,686	360	7.46 ***
Fair	958	259	3.70 ***
Good	547	218	2.51 **
Very good	311	214	1.45
Functional Status (no ADLs omitted)			
5-6 ADLs	1,687	386	4.37 ***
3-4 ADLs	860	328	2.62 ***
1-2 ADLs	510	210	2.43 **
Difficulty walking 2-3 blocks ("no" omitted)			
Unable/ a lot of difficulty	1,567	261	6.01 ***
Some/ a little difficulty	471	201	2.34 **
Difficulty lifting ("no" omitted)			
Unable/ a lot of difficulty	865	249	3.47 ***
Some/ a little difficulty	392	190	2.06 **
Chronic Conditions			
Heart Attack	1,169	244	4.79 ***
Angina	657	240	2.74 ***
Other Heart Conditions	571	183	3.11 ***
Cancer, except skin	457	201	2.27 **
Diabetes	1,476	217	6.80 ***
COPD	935	232	4.04 ***
R Squared	-		
Adjusted R Squared	-		
Observations	27,130		
F, Chi-Squared	359.2 ***		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor.
A variance components model was estimated using SAS PROC Mixed, which does not produce R-squares.
Identical to Model (4) in Table 6-2. Coefficients and standard errors are in 1992 dollars.

*** Statistically significant at 1 percent level.

** Statistically significant at 5 percent level.

* Statistically significant at 10 percent level.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), 7(1993) and 10(1994).

6.3.5 Model with Selected Chronic Conditions, Disabled Only

The fifth model we present includes the same set of variables as our preferred elderly model (Model (4)), but is estimated for the much smaller sample of disabled respondents (n = 5,205). Unfortunately, this model exhibits undesirable instabilities. Several coefficients are negative and statistically insignificant. Hence, to estimate a stable model for the disabled, a reduced set of variables must be employed.

6.3.6 Model with Collapsed Health Status Variables, Disabled Only

Model (6) is based on Model (5), but multi-level health status variables are collapsed to dichotomous variables. The following distinctions are made:

- self-rated health status: poor/fair versus good/very good/excellent;
- functional status: 5-6 ADLs versus 0-4 ADLs; and
- walking or lifting: unable to do/a lot of difficulty versus some/a little/no difficulty.

In addition, "heart attack", "angina", and "other heart conditions" are combined into a single "heart disease" chronic condition. "Cancer" is dropped because it is statistically insignificant for the disabled.

This model has plausible and statistically significant coefficients. It achieves an adjusted R-square of 3.4 percent. The survey measures do not predict payment variation as well among the disabled as the elderly (i.e., the R-square is lower) because levels of measured disability and poor health are high in the disabled population, so there is less variation on the health status measures among the disabled than the elderly. Note that the intercept in Model

(6) is $1,351, which is the implied payment level for a female reporting no chronic conditions, excellent self-rated health, and no disabilities. This is in contrast to a negative predicted payment for the same individual in the combined aged/disabled model (Models (1) and (2)), indicating the need for a separate model for the disabled. Better models for the disabled could be developed with larger sample sizes, but this type of "collapsed" model is probably the best for which stable coefficient estimates can be obtained with currently available sample sizes.

6.3.7 Model with Collapsed Health Status Variables, Disabled Only, Estimated with Variance Components

Model (7) is identical to Model (6) except that the variance components technique is used to estimate it. Again standard errors rise, reducing statistical significance. Coefficient estimates are more unstable than with the elderly sample because of the smaller number of observations. Model (7) is our preferred model for the disabled, although we feel that survey risk adjustment models for the disabled need more refinement before they are ready for use in payment.

6.3.8 Conclusions

In this chapter we have used three pairs of years of MCBS survey data to obtain a larger sample size and achieve more stable estimates of a preferred survey risk adjustment model. Our preferred model was developed to have good ability to predict future Medicare payments, to utilize only variables appropriate for payment purposes, and to be implementable

from responses to the HoS survey. Our final model for the aged (Table 6-3) consists of age, sex, self-rated health status, functional status measured by ADL count, physical functioning measured by difficulty in walking and lifting, and six chronic conditions: heart attack, angina, other heart conditions, cancer, diabetes, and COPD. Although we believe our estimated parameters will be reasonably stable in other samples, many of the coefficient standard errors are still large. Thus, the model would benefit from being estimated on even larger sample sizes in future work. In addition, it would be useful to validate the final model in future work, as was done for other models in Chapter 5.

We were less successful in developing a preferred survey model for the under-age-65 Medicare disabled population. Claims-based models seem superior for this population (see Chapter 5). Relatively small sample sizes of disabled limited the utility of our analysis. We recommend further study of survey risk adjustment for the disabled before consideration of these models for payment adjustments.

7

Conclusions

No one risk adjustment model is best on all empirical criteria considered in this report. The claims-diagnosis-based DCG/HCC model has greater overall predictive power than the survey, PIPDCG, and prior use models, and predicts average expenditures as or more accurately for most of the validation subgroups we considered. It appears to be the best single model empirically. However, for certain subgroups -- the elderly receiving help with activities of daily living, for example, -- it doesn't appear to predict expenditures as accurately as certain of the survey models. No model predicts uniformly well for all groups. Thus, which model is preferred depends in part on what relative weight policymakers put on "getting payment right" for different subgroups of Medicare beneficiaries.

Practical and administrative considerations are also important in evaluating alternative claims and survey adjusters. The DCG/HCC model requires ambulatory diagnoses and hence encounter data systems, which are expensive and time-consuming to develop, although useful for a variety of purposes once implemented. The PIPDCG model requires only principal inpatient diagnoses, which are typically much less expensive to obtain than ambulatory diagnoses, of greater clinical validity, and easier to audit. Claims adjusters are sensitive to intentional and unintentional variations in diagnostic coding (e.g., "upcoding"). The PIPDCG model is much less sensitive to the completeness of diagnostic coding than the DCG/HCC

model, but it is sensitive to which inpatient diagnoses are "principal" versus secondary. Moreover, a beneficiary must be hospitalized for his or her diagnosis to be taken into account in the PIPDCG model. This penalizes efficient health plans that avoid hospitalizations and sets up possibly inappropriate incentives for hospitalization.

Surveys have lower startup costs and are available more immediately than claims diagnoses, but are expensive and burdensome to conduct on an ongoing basis.[1] They suffer from nonresponse, and biased and inaccurate responses (e.g., what does self-rated health mean from someone with dementia?). Providers may be able to influence survey responses (e.g., by "prescribing" disability) and beneficiaries may respond strategically once they realize that provider reimbursement depends on their survey answers. Survey responses may deviate from "objective" criteria along socio-demographic or regional lines, and are difficult to audit or verify. Either claims or survey adjusters would need to be recalibrated to adjust for behavioral changes in diagnostic coding or survey responses if either were implemented for payment.

Adding survey variables to a claims-based model such as the DCG/HCC increases overall explanatory power and improves predictions for key subgroups such as the elderly receiving help and dual Medicaid/Medicare eligibles. But a combined model requires obtaining both survey and encounter data, which might be prohibitively expensive. Also, overpredictions for certain diagnoses (osteoporosis, hip fracture, dementia) are increased. Combined models warrant more research as a means of combining diagnoses from claims with

[1] The marginal cost of using surveys for risk adjustment may be low if the same survey instrument (e.g., the SF-36) that is used for outcomes assessment can also be used for risk adjustment (perhaps with a few added questions).

severity/disability information (subjective health, functional status) from survey responses into a single powerful model.

Substantial redundancy exists among the various survey adjusters. Their combined explanatory power is much less than the sum of their individual explanatory power. Nevertheless, independent dimensions of health status are measured by the different survey variables. A multi-dimensional survey model like our preferred survey model (see Chapter 6) or the SF-36 simulation is necessary to predict expenditures well across the range of subgroups. The disadvantage of multi-dimensional survey models is that survey instruments must be longer, increasing survey expense and respondent burden, and lowering response rates.

Although multiple domains of health status need to be surveyed, redundancy implies that some pruning of questions based on other criteria is possible and desirable. Other desirable characteristics for risk adjusters include resistance to manipulation by providers or beneficiaries, objectivity, reliability, parsimony, and face validity. In our opinion, certain survey variables rank higher on these criteria than others. We would place chronic conditions (diabetes, heart disease), physical functioning ("can you walk two blocks?"), and activities of daily living ("do you have difficulty bathing") higher on this scale and social functioning ("has your health interfered with your social activities?"), self-rated health ("is your health excellent, good, fair, or poor?") and instrumental activities of daily living ("do you have difficulty preparing meals?") lower. Others might disagree with our assessment. More research and

practical experience is needed on pertinent aspects of survey adjusters other than predictive power.

Even the best survey models don't predict accurately for groups defined on prior medical expenditures. Providers will be able to practice substantial risk selection against survey models by employing their knowledge of the medical care use of actual or potential enrollees. Although prior use and claims-diagnosis models worked well, our survey models also did not perform well for the disabled. At a minimum, parameters for the elderly and the disabled appear to be different and require separate estimates. Perhaps totally different survey models need to be developed for the disabled.

A final, and very important point, is that more data are needed to obtain stable and reliable estimates of risk adjustment models. Although we believe that the combined year estimates of our preferred survey risk adjustment model (Chapter 6) are plausible and reasonably stable, they are less precise than would be ideal (as indicated by the relatively large standard errors of estimates). Not surprisingly, comparison of parameter estimates on single-year 1991/1992 versus 1992/1993 data shows substantial differences (Chapter 4). For example, "osteoporosis", which has a highly statistically significant coefficient of $1,312 in the 1991/1992 chronic conditions model has a negative and statistically insignificant coefficient of -$288 in the 1992/1993 model. "Very good" in the self-rated health status model has a statistically insignificant coefficient of $486 in 1991/1992 versus a highly significant coefficient of $1,003 in 1992/1993. More data will also increase the sensitivity of

risk adjustment models by allowing the estimation of health status scales with more response levels (e.g., "a lot/some/a little/no difficulty" versus "some/no difficulty").

7.1 Limitations

This study has several significant limitations. We analyzed a particular set of survey models, albeit a wide range of this class of models. We analyzed only two claims-diagnosis-based models, the DCG/HCC and PIPDCG models, not, for example, the Ambulatory Care Groups (ACG) model (Weiner *et al.*, 1996; Weiner *et al.*, 1991). Our results may not generalize beyond the particular survey and claims-based models we analyzed. Nor will our results necessarily generalize to other populations. For example, the differences between our results and those of Fowles *et al.*, 1996 may be attributable to that study's evaluation of the actual SF-36 survey scales and the Ambulatory Care Groups claims model on a mixed under-65 and aged population.

Because many high cost medical conditions that account for a large portion of expenditures are rare, large sample sizes, such as are available from claims files, are desirable for estimating and validating risk adjustment models. Sample sizes comparable to claims samples are not currently available for survey variables. Our MCBS results--both estimation and validation--may not fully generalize to other samples because of the MCBS's limited sample size. That is, our results are influenced to some extent by random error. Nevertheless, we believe that most of our qualitative findings will generalize to other and larger samples.

Another technical limitation is that we did not have a fully independent validation sample, which may tend to overstate the predictive power of all risk adjustment models. We did not include individuals in our sample for whom we did not have survey responses, either because a person did not respond to the MCBS at all, or because he or she did not answer a specific question. To the extent that survey nonrespondents are sicker on average than respondents, our results may somewhat overstate the predictive power of models (especially survey models) compared to a full set of responses.

References

Calkins D, Rubenstein L, Cleary PD *et al.*: Failure of physicians to recognize functional disability in ambulatory patients. *Ann Intern Med* 114: 451-4, 1991.

Daley JD, Jencks S, Draper D *et al.*: Predicting hospital-based mortality for medicare patients: a method for predicting mortality for patients with stroke, pneumonia, acute myocardial infarction, and congestive heart failure. *JAMA* 260:3617-24, 1988.

Elam JT, Graney MJ, Beaver T *et al.*: Comparison of subjective ratings of function with observed functional ability of frail older persons. *American Journal of Public Health* 81: 1127-1130, 1991.

Ellis RP, GC Pope, LI Iezzoni *et al.*: *Diagnostic Cost Group (DCG) and Hierarchical Coexisting Conditions (HCC) Models for Medicare Risk Adjustment.* Final Report, HCFA Contract No. 500-92-0020, April 1996.

Ellis, RP, Ash, A: Refinements to the Diagnostic Cost Group Model. *Inquiry* 32:1-12, 1995.

Ellis, RP, Pope, GC, Iezzoni, LI, *et al.*: Diagnosis-Based Risk Adjustment for Medicare Capitation Payments. *Health Care Financing Review* 17:101-128, 1996.

Fowles, JB, Weiner, JP, Knutson, D, *et al.*: Taking Health Status Into Account When Setting Capitation Rates: A Comparison of Risk-Adjustment Methods. *Journal of the American Medical Association* 276:1316-1321, 1996.

Fowles, JB, Weiner, JP, Knutson, D: *A Comparison of Alternative Approaches to Risk Measuresment: Final Report to Physician Payment Review Commission.* Minneapolis, Minnesota: Park Nicollet Medical Foundation. Nicollet Medical Foundation Grant 93-G07, 1994.

Gruenberg, L, Kaganova, E, Hornbrook, M: Improving the AAPCC With Health Status Measures From the MCBS. *Health Care Financing Review* 17:59-76, 1996.

Guccione AA, Felson DT, Anderson JJ *et al.*: The Effects of Specific Medical Conditions on the Functional Limitations of Elders in the Framingham Study. *American Journal of Public Health* 84(3): 351-358, March 1994.

HCFA Statistics, Table 6, pg 8, HCFA Publication #03403, October 1997.

Hornbrook, MC, Goodman, MJ: Chronic Disease, Functional Health Status, and Demographics: A Multi-Dimensional Approach to Risk Adjustment. *Health Services Research* 31:3, August 1996.

Hornbrook, MC, Goodman, MJ: Assessing Relative Health Plan Risk with the RAND-36 Health Survey. *Inquiry* 32:56-74, 1995.

Iezzoni LI: Risk Adjustment for Medical Effectiveness Research: An Overview of Conceptual and Methodological Considerations. *Journal of Investigative Medicine* 43(2): 136-150, April 1995.

Iezzoni LI, Heeren T, Foley SM *et al.*: Chronic conditions and risk of in-hospital death. *Health Serv Res* 29: 435-60, 1994.

Keeler EB, Kahn KL, Draper D *et al.*: Changes in sickness at admission following the introduction of the prospective payment system. *JAMA* 264: 1962-8, 1990.

Kronick, R, Dreyfus, T, Lee, L, Zhou, Z: Diagnostic Risk Adjustment for Medicaid: The Disability Payment System. *Health Care Financing Review* 17:7-34, 1996.

Manning, WG, Newhouse, JP, Ware, JE: The Status of Health in Demand Estimation, or Beyond Excellent, Good, Fair, Poor. In V.R. Fuchs ed. *Economic Aspects of Health.* Chicago: University of Chicago Press, pp. 143-184, 1982.

Mor V, Wilcox V, Rakowski W and Hiris J: Functional Transitions among the Elderly: Patterns, Predictors, and Related Hospital Use. *American Journal of Public Health* 84(8): 1274-1280, August 1994.

Pope, G. *et al.*, Revised Diagnostic Cost Group/Hierarchical Coexisting Condition Models for Medicare Risk Adjustment, Final Report to the Health Care Financing Administration, February, 1998. Waltham, MA: Health Economics Research, Inc.

Pope, G.C., Adamache, K.A., Khandker, R., Walsh, E.: *Evaluating Alternative Risk Adjusters for Medicare.* Draft Report to the U.S. Health Care Financing Administration. (Waltham, MA: Center for Health Economics Research), 1997.

Rubenstein LV, Calkins DR, Greenfield S *et al.*: Health status assessment for elderly patients: report of the society of general internal medicine task force on health assessment. *J Am Geriatric Soc* 37: 562-9, 1988.

Rudberg MA, Parzen MI, Leonard LA and Cassel CK: Funtional Limitation Pathways and Transitions in Community-Dwelling Older Persons. *The Gerontologist* 36(4):430-440, 1996.

Spector WD: Functional Disability Scales in Quality of Life and Pharmacoeconomics in Clinical Trials, second edition, edited by B. Spilker. Lippincott-raven Publishers, Philadelphia, 1996.

Verbrugge L, Merrill, S and Liu, X: Measuring Disability with Parsimony. Proceedings of the 1995 Public Health Conference on Records and Statistics. Hyattsville, MD: National Center for Health Statistics.

Verbrugge LM, Lepkowski JM and Imanaka Y: Comorbidity and Its Impact on Disability. *The Milbank Quarterly* 67(3-4): 450-484, 1989.

Ware, J.E.: SF-36 Health Survey Manual and Interpretation Guide. Boston, MA: The Health Institute, New England Medical Center, 1993.

Ware, JE, Sherbourne, CD: The MOS 36-Item Short-form Health Survey (SF-36). *Medical Care* 30:473-483A, 1992.

Weinberger M, Samsa GP, Schmader K *et al.*: Comparing proxy and patients' perceptions of patients' functional status: Results from an outpatient geriatric clinic. *Journal of American Geriatric Society* 40:585-588, 1992.

Weiner, JP, Dobson, A, Maxwell, S, *et al.*: Risk-Adjusted Medicare Capitation Rates Using Ambulatory and Inpatient Diagnoses. *Health Care Financing Review* 17:77-100, 1996.

Weiner, JP, Starfield, B, Steinwachs, D, Mumford, L: Development and Application of a Population Oriented Measure of Ambulatory Care Case-Mix. *Medical Care* 29:452-472, 1991.

Wolinsky FD, Callahan CM, Fitzgerald JF and Johnson RJ: Changes in Functional Status and the Risks of Subsequent Nursing Home Placement and Death. *Journal of Gerontology* 48(3): S93-S101, 1993.

Appendix A

Alternative Disability Models

A Alternative Disability Models

Survey-based approaches to risk adjustment offer the opportunity to include functional status measures that are not available in claims data. The MCBS includes several domains of functional status, and within each domain, there are several options for variable definition. To develop survey-based risk adjustment models based on the MCBS, we had to determine which of the available functional status items to include and how to define the selected variables. This appendix details the theoretical issues and our empirical work analyzing alternative functional status measures in the MCBS. We used these exploratory analyses as a basis for selecting measures to include in the individual and comprehensive survey-based models presented in Chapter 4 and for our preferred survey model, presented in Chapter 6.

In this appendix, we present a detailed discussion of functional status measures in relation to health and healthcare costs. The first section is a brief review of the literature relating functional status to morbidity, mortality and utilization of medical care. The second section describes the functional status measures available in the MCBS. Finally, the last section provides detailed analyses of the self-reported functional status measures available in the MCBS and their relation to Medicare payments.

A.1 Literature Review

In this report we are concerned with a very narrow aspect of components of disability, those related to medical care costs. While functional impairments and disability are more often considered in relation to long-term care utilization and costs, functional status can be related to Medicare costs in several ways:

- as measures of disease severity, impact, and acuity
- as the result of comorbidities or interactions among medical conditions
- as risk factors for further deconditioning and medical sequelae
- as indicators of as yet undiagnosed medical conditions.

The increased medical care utilization associated with functional impairment relates to:

- treatment of the underlying causes of impairment (e.g., joint replacement or post-stroke rehabilitation)
- prolonged hospitalization, SNF and home health utilization after acute episodes
- ongoing utilization related to the functional impairments.

Functional Status and Medical Costs: Iezzoni (1995) refers to functional status as representing "a final common pathway of disease progression." As such, functional status reflects the impact of individual and combinations of medical conditions (Guccione *et al.*, 1994; Verbrugge *et al.*, 1989), both chronic and acute, on the individual. Not only are functional status measures a type of severity indicator for illness or injury, but impaired

functional status is also a risk factor for additional morbidity, mortality, and health service utilization.

The potential value of functional status measures to risk adjustment is also related to their appropriateness as self-reported measures. While recall of medical information may be imperfect, people are able to report accurately their own functional status. Indeed, discrepancies found between self-report and physician report have been shown to reflect physicians' underestimation of disability (Calkins *et al.*, 1991). In addition, individuals who may have as yet undiagnosed conditions, are able to notice and report functional difficulty. There are also limitations to functional status measures as predictors of acute care utilization. Functional impairments resulting from musculoskeletal problems, in the absence of comorbid conditions, are not necessarily associated with high medical costs. In addition, temporary changes in functional status related to acute conditions may not predict future utilization.

While self-reported functional status is correlated with self-rated health status, it is distinct and different in its relation to morbidity, mortality and utilization.

Relation to Morbidity: A strong association is noted between the number of chronic conditions and lower levels of function (Iezzoni, 1995; Verbrugge *et al.*, 1989). This association is advantageous to the use of survey measures for risk adjustment. Functional status information can proxy for some degree of incomplete medical diagnosis information. Self-report of functional impairment may also precede diagnosis and evaluation related to the impairments (Verbrugge, *et al.*, 1989; Hornbrook and Goodman, 1995) either as early

symptoms or as evidence of illness in persons with limited inclination or access to seek medical care.

Several investigators have studied the relationship between specific medical conditions and functional impairment. Guccione and his colleagues used elderly sample members in the Framingham Study to evaluate the relation of specific medical conditions to disability (1994). In their study, osteoarthritis of the knee, heart disease, depressive symptomology, and stroke created the highest levels of disability. The disability measures they used were a combination of physical activities and IADLs, including stair climbing, walking a mile, heavy housework, housekeeping, cooking, grocery shopping and carrying bundles. Using the Study on Aging (SOA) Verbrugge, Lepkowski and Imanaka (1989) studied the relation of medical conditions to a series of disability measures, individually and in pairs. The disability measures they included were difficulty walking; difficulty with a series of physical functions including walking a quarter of a mile, walking up ten steps, standing 2 hours, sitting 2 hours, stooping/crouching/kneeling, reaching up over head, reaching out, grasp, lifting/carrying 25 pounds, and lifting/carrying 10 pounds; five ADLs; 5 IADLS; and a measure of global role limitation. They found wide variation in the relation of both individual conditions and pairs of condition to disability. For example, Verbrugge and her colleagues found cardiovascular disease, hip fracture, visual impairment, osteoporosis and atherosclerosis to have the greatest individual impact on disability. Combinations of these diseases and specific combinations of some of these conditions and diabetes or ischemic heart disease were also significantly related to the series of disability measures. However, hearing impairment, arthritis and hypertension,

the most prevalent conditions, were not generally associated with high levels of disability individually or in combination with each other. They did find hypertension in combination with other circulatory problems, and ischemic heart disease in combination with cancer, to be significantly associated with most disability measures.

These findings relating medical conditions, singly and in combination, to disability measures are significant for risk adjustment in several ways. First, they suggest that counts of medical conditions alone do not effectively capture the cost effects of the less prevalent but more disabling conditions. Risk adjustment methodologies based on either individual conditions or counts also miss the synergistic effects of particular comorbidities. Finally, conditions may be as yet undiagnosed, but individuals will notice changes in their functional status and be able to report such changes in surveys. Hence the relation between conditions and disability allow the use of disability measures as a proxy for some medical conditions individually or in combination.

Changes in Functional Status: While functional status generally remains stable over time or declines, some decreased functioning is transient, indicating a relation to acute as well as chronic conditions. Mor and his colleagues (1994) report about 12 percent of elders in a longitudinal study based on the Longitudinal Study of Aging (LSOA) had fewer IADL impairments over a 6 year period than at baseline. Mor also found that 3.5 percent of women and 7 percent of men age 70-79 who were in the most severely impaired group (3 or more ADL impairments) were completely independent 6 years later. These findings suggest that some impairment is associated with acute conditions and hence is a marker of short term

morbidity- for example after a heart attack, during cancer treatment or before post-stroke recovery or rehabilitation. Alternatively, functional impairments may also indicate a need for medical treatment, for example, joint replacement surgery and subsequent rehabilitation. In this case, functional impairment may predict short term costs, but not be good predictors of long-term costs. In addition, improved functional status as a result of rehabilitation or recovery decreases a person's risk of medical sequelae related to functional impairment.

It is important, however, to note that the majority of those with functional impairment remained stable or declined overtime. Indeed, Mor and his colleagues found that functional status at baseline, along with age, self-rated health and comorbidity predicted functional status at the 6-year follow-up.

Relation to Mortality: Medical costs are well known to be high in the months preceding death. Several studies have linked functional status impairments to increased risk of mortality. Using the Established Populations for Epidemiologic Studies of the Elderly data (EPESES) and a multivariate analysis, Corti *et al.*, (1994) found increased relative risks associated with mobility impairment of 1.9 for men and 1 7 for women. They found even greater relative risk for persons with at least one ADL impairment: 2.6 for men and 2.5 for women. A study of Medicare quality by RAND found increased risk of death for Medicare beneficiaries hospitalized for pneumonia who were unable to walk prior to hospitalization. Similarly, Daley *et al.*, (1988), found the need for assistance with walking or inability to walk prior to hospitalization a significant predictor of 30 day mortality for Medicare beneficiaries hospitalized for pneumonia or congestive heart failure. Iezzoni *et al.*, 1994, report an odds

ratio for inhospital death of 1.54 for the 3.7 percent of persons in their sample identified as functionally impaired. These findings clearly link functional status to increased risk of mortality.

Utilization: Functional impairment is associated with higher hospitalization rates (Mor *et al.*, 1994), longer hospital stays (Iezzoni, 1995) and increased utilization of skilled nursing facility care, and home health care in both acute and chronic situations. Mor *et al.* (1994), also found that those with declining functional status over time had the highest hospital utilization rates, those remaining dependent over time the next highest, and those whose functional status improved over time had the lowest subsequent hospital utilization rates. Harris, *et al.* (1989) study of the oldest old in the Longitudinal Study of Aging focused on the one third of those 80 years of age and older who reported no problems with walking, lifting 10 pounds, climbing 10 steps or with stooping, crouching, or kneeling. They reported much lower relative odds of hospitalization (0.4), physicians (0.6) and nursing homes (0.3) in survivors over a two year period for this group with no physical impairments. Hence the absence of functional impairment, or improvement in functional status, is associated with lower health care costs.

A.2 MCBS Disability Measures

A.2.1 Overview

The MCBS contains several types of disability measures including functional status measures and functional impairment measures. These measures include activities of daily

living (ADLs), instrumental activities of daily living (IADLs), physical impairment measures and the effect of health on social activities. In this section we review the disability measures and their construction in the MCBS. Descriptive statistics and the results from using the different measures as predictors of future costs are presented in the next section. Finally, we present our conclusions regarding the relative merits of the disability measures as predictors of Medicare expenditures.

Functional Status: Functional status generally refers to the degree to which a person can perform basic tasks of daily life and is generally categorized into ADLs and IADLs. To the extent a person cannot perform basic tasks, s/he is categorized as disabled. The ADLs included in the MCBS are bathing, dressing, mobility, transferring, toileting and eating. The ADLs included and their wording reflect a variation of the Katz ADL scale, with the addition of mobility and exclusion of incontinence. We chose to include the urinary incontinence variable available in the MCBS along with the chronic conditions as it is often considered an physical impairment but not a disability. The IADLs in the MCBS are telephone use, shopping for personal items, money management, light housework, heavy housework and meal preparation. Functional status measures are associated with chronic disability, disease severity, and acute phases of health conditions such as recent stroke or myocardial infarction.

Physical Impairments: Physical impairments refer to the effect of organ system problems on physical function, but do not necessarily result in disability, i.e., the inability to perform necessary tasks. The physical impairment measures included in the MCBS are degree of difficulty walking 2-3 blocks; lifting or carrying 10 pounds; stooping, bending or kneeling;

reaching over one's head; and writing. Like the functional status measures, physical impairments are indicators of relative severity or the interactions of comorbid conditions. However, they may have additional value as markers of as yet undiagnosed health problems.

Social Function: The effect of health on social activities is another dimension of disability available in the MCBS. Impact of health on social function is a common variable in disability surveys. Verbrugge, Merrill and Liu (1995) suggest it is a global measure of disability, much as self-rated health is a global measure of health.

A.2.2 Variable Construction

The MCBS is similar to other surveys that include a series of questions regarding each ADL. The series asks first if a person has difficulty with the task due to a health reason, no difficulty, or if the person does not perform the task. Those who do not perform the tasks are asked if this is due to a health reason. Those who report difficulty are asked if they receive any help. Those who receive help are asked about the type of help they receive. This series allows impairment to be defined by presence of difficulty, receipt of help or type help received (hands-on, or supervision). We created two dichotomous variables from this series. The first, "difficulty" is the more inclusive variable. "Receipt of help" is a subset of the group reporting difficulty.

Difficulty = 1, if reports difficulty or nonperformance due to a health reason

= 0, if reports no difficulty or nonperformance without citing health reason;

Receives Help = 1, if reports hands-on help or supervision by another person

= 0, if does not receive hands-on help or supervision.

We examine both to observe the effects of the broader, with difficulty, measure and the more targeted, receipt of help measure. Table A-1 displays the percent reporting difficulty or help for each individual ADL in 1991 and in 1992.

As shown in this table, reports of difficulty are more frequent than receipt of help in each task. The pattern across the two years reported remains consistent, although fewer individuals reported difficulty or help in 1992 than in 1991. This difference across the two years might be due to the composition of the replacement sample in 1992. Attriters were replaced with younger Medicare beneficiaries.

Instrumental Activities of Daily Living. All community residing respondents were asked about both difficulty and receipt of help with each IADL task in a manner similar to that used for the ADLs. Among the institutionalized respondents, data were only collected for telephone use, money management and shopping for personal items because light housework, heavy housework and meals are provided to all institutional residents. As a result, we imputed values of both difficulty and help in all three of these tasks.

Table A-1 displays the percent reporting difficulty or receipt of help in both years. IADL impairments at both levels, difficulty and with help, were more frequent than ADL impairments. As with the ADLs, only a subset of all persons reporting difficulty also report receipt of help, and the patterns are similar across the years. However, many more of those reporting difficulty also reported help with IADLs than with ADLs.

Functional Status Scales. ADLs and IADLs are often used to create scales. We created scales based on the number of ADL impairments, the presence of IADLs only, or the absence of any functional impairments. This is a conceptually straightforward approach: those with a greater number of ADL and IADL impairments are more disabled than those with fewer. In addition, research related to the Katz ADL scale, upon which the MCBS measures are based, has shown a progression of impairments. We created scales based on report of difficulty and receipt of help as for the individual tasks s/he receive help with the task. Table A-2 displays these scales and the percent reporting difficulty or help in each year. As with the individual ADLs and IADLs, the number reporting difficulty greatly exceed those receiving help, and the distributions are similar in 1991 and 1992. There is a lower level of disability reported overall in 1992 than in 1991. Only about 20-25 percent report receipt of help with ADLs, while about 40-50 percent report difficulties in ADLs.

Physical Impairments. After analyzing the detailed responses for each physical impairment, we collapsed the responses to a dichotomous variable indicating no difficulty or any degree of difficulty with each activity. As shown in Table A-3, about 70 percent of the sample report trouble stooping, kneeling or bending, about half report difficulty walking 2-3 blocks, and those reporting difficulty lifting approach 50 percent. While there is some variation from 1991 to 1992, the pattern is similar.

Social Function. The MCBS asks "How much of the time in the past month has (your/SP's)[1] health limited (your/SP's) social activities, like visiting with close friends or close

[1] SP refers to subject person when a proxy responds on behalf of the beneficiary.

relatives?" The majority, 62.5 percent, of the sample responded none of the time, 7.4 percent responded all of the time, 30 percent responded some or most of the time. The community dwelling respondents are given a 4 point scale for responses ranging from none of the time to all of time. The institutional respondents have a 6 point set of possible responses. We collapsed the options in the institutional responses to a 4 point scale. We also created a dichotomous variable separating those who reported any degree of social function limitation from those who reported no social function limitations.

A.3 Disability Models

A.3.1 Overview

In this section, we present the results of using several models of disability as predictors of Medicare costs in the next year. All models add disability measures to the same set of age categories and the gender variable used in all of our models (see Chapter 4). We present models based on individual functional status measures, counts of functional status limitations, physical limitations and social function models. The functional status models include those based on report of difficulty and those based on receipt of help. All models are reported for both 1991 characteristics as predictors of 1992 costs and 1992 characteristics as predictors of 1993 costs.

In reviewing these findings, it is important to keep several factors in mind. First, while the adjusted R-squares appear low (ranging from 0.016 to 0.034) these represent substantial

improvements over the predictive power of the AAPCC. Second, the coefficients, representing dollar differences in predicted Medicare expenditures, can be substantial.

Before proceeding to the detailed descriptions of the individual models, we present a summary table of the adjusted R-squares (Table A-4). Key findings from this table include:

- The highest adjusted R-squares are associated with the models combining all three types of variables: ADL/IADLs, physical impairments, and social function.
- While there is a two-fold difference between the lowest and highest adjusted R-squares, models incorporating only one type of disability measure (either ADLs/IADLs, Physical Impairments or Social Function) all explain about 2 – 2 ½ percent of the observed variation in Medicare expenditures.
- Combining the measures increases the predictive power to about 3%. This suggests substantial redundancy among these measures since their combined predictive power is much less than the sum of their individual explanatory power.
- There is not much difference in adjusted R-squares for ADL and IADL models between those defining impairment as "reports difficulty" and those based on "receipt of help" (with the exception of lower adjusted R-square for the receipt of help model in 1991).
- Models using ADL/IADL scales have almost as much predictive power as those incorporating each ADL and IADL separately.

A.3.2 Functional Status Models

Individual ADLs and IADLs. These models, displayed in Table A-5, are the most disaggregated of the functional status models. We present models based on the difficulty cut-off and on the receipt of help as the definition of impairment, identifying the coefficients and statistical significance of each as well as the R-square and adjusted R-square for each model.

Table A-6 identifies the statistically significant variables across the models. Across the models, ADL problems with mobility and bathing have positive and significant coefficients, as does heavy housework among the IADLs. Help with money management has a negative coefficient in both years. The negative coefficients associated with money management and telephone use may be due to their relation to dementia, which we also found to be associated with lower costs in the chronic condition models.

The instability of individual coefficients across the two years, some of which change sign as well as magnitude and statistical significance, is a major factor in our recommendation to use a scale rather than individual impairments in functional status.

Individual ADLs. We also tested models based on the individual ADLs only, alternatively using difficulty as the cut-off or help, and for each of the two years' data. The data are presented in Table A-7. Interestingly, age 0-64 loses significance in the receipt of help models. In other words, receipt of help with ADL tasks appears to be strongly associated with Medicare costs in the younger disabled population. In most of our models, this age category also has a negative and significant coefficient. In the difficulty models, this age category has a negative and significant coefficient in 1991/92 and a negative and significant coefficient at the .10 level in the 1992/93 model, because the under age 65 are by definition disabled (i.e., have many ADL and IADL impairments) but are not more expensive than the elderly on average. Mobility and bathing have sizable and significant coefficients in all four models.

There are some interesting differences between this model and the previous model that includes individual IADL impairments. The most noticeable difference is in the bathing coefficients, which almost double in every individual ADL model. Dressing was insignificant in the ADL/IADL model, but significant in this model, suggesting that there is overlap between dressing and IADL impairments. Finally, the adjusted R-squares of the ADLs only models are only slightly lower (about 15 percent) than those of the ADL/IADL model, suggesting considerable redundancy between the IADL and ADL measures.

Functional Status Scales. Table A-8 presents the models based on a functional status scale: 5-6 ADLs, 3-4 ADLs, 1-2 ADLs, IADLs only, and no impairments. Again, we tested models based on difficulty or receipt of help in both years. The scale based on report of difficulty is the model we selected to report in Chapter 4 and validate in Chapter 5. Both variants follow the same basic pattern: generally greater numbers of impairments are highly significant and are usually associated with higher Medicare expenditures. The models based on report of difficulty follow the pattern precisely, with step-like increments as the number of impairments increase. The models based on receipt of help follow the general pattern, but there is less variation in the size of the coefficients and in the second year, the group with 5-6 ADL impairments has a lower coefficient than that with 3-4 ADL impairments. As expected the coefficients of the scales are more stable across years than the coefficients of the individual ADLs or IADls.

A.3.3 Physical Impairment Models

We present the physical impairment model estimated on the 1991/92 data and the 1992/93 data in Table A-9. Difficulty walking 2-3 blocks and difficulty lifting or carrying 10 pounds are highly significant and have sizable coefficients in both years. The significance of age, as in some of the other disability models, is reduced, although age less than 65 remains significant and negative. The adjusted R-square is 0.0192. Thus, a model only asking two disability questions, "Do you have any difficulty walking 2-3 blocks?" and "Do you have any difficulty lifting or carrying 10 pounds?" has almost as much predictive power as a model requiring asking a series of ADL questions. The physical impairments model offers a parsimonious approach to disability measurement.

A.3.4 Social Function

We present two sets of regressions relating the effect of social function on Medicare payments. Table A-10 presents models with the four point scale set of responses specified individually. Those who reported no impairment in social function related to health are the omitted group. Table A-11 presents models in which we collapsed the responses to a dichotomous variable indicating any degree of social function impairment. The adjusted R-square of the expanded 1992/93 model, 0.0274, is essentially equivalent to the fully specified individual ADL and IADL model. The adjusted R-square of the expanded 1991/92 model, 0.0191 is not much lower than the fully-specified individual ADL and IADL model for that year. Hence, with one simple question, almost as much variation is explained as with

a series of questions. The second model, in which all degrees of social function impairment are collapsed into one, yields adjusted R-squares of 0.0143 in the 1991/92 model and 0.0198 in the 1992/93 model.

It is also interesting to compare the coefficients that result from the social function model to those in the ADL models. The coefficients show similar patterns and magnitude for both measures of severity of impairment. Those who report that health interferes with their social activities all of the time have a coefficient over $5,000 in both years. This is similar to the coefficient for those reporting 5-6 ADL impairments. Yet correlations among the variables (not shown in this report) indicate only partial overlap between the level of social function impairment and ADL/IADL impairment. It appears that the highly significant coefficients for the age variables in the social function models, increasing in magnitude with age, pick up increasing ADL/IADL impairment. As we saw in earlier sections, age 85 and over became insignificant with increasing ADL/IADL detail. It is also interesting to note that some of the social function models are the only models in which gender loses significance. This indicates a correlation between gender and social function.

A.3.5 Combined Models

In this section we present the results of two models combining the disability variables. We entered all the disability measures together in backward stepwise regressions to identify the variables most responsible for the variability in Medicare payments.

Combining ADLs and IADLs, Physical Impairments, and Social Function Limitations. Our final analyses regress Medicare payments on all three domains of functional status measurement combined. We employed backwards stepwise regression. This form of regression analysis first enters all the variables and then drops those variables which contribute least to the R-square. The procedure stops when dropping additional variables would decrease the R-square meaningfully. We forced the age variables and gender to stay in the regression, even if nonsignificant.

We ran this procedure for both years using both individual ADLs and IADLs and the ADL/IADL scale. Table A-12 reports the findings for the models incorporating individual ADLs and IADLs, based on report of difficulty, with the dichotomous physical impairment and social function limitation measures. In both samples, 1991/1992 and 1992/1993, bathing and walking are the only individual ADLs that remain in the model. Of the IADLs, meal preparation remains in both samples and has a highly significant coefficient of about $1,300. While heavy housework is significant in the first sample, it drops out in the second sample. Ability to use the telephone has a large negative coefficient which is only significant in the second sample. Social function limitation retains sizable and highly significant coefficients across both samples: $922 in the first sample and $1,277 in the second.

This is the first model in which we observe instability in the physical impairment variable "walking 2-3 blocks." While it is sizable ($864) and significant in the first sample, it drops out in the second. Difficulty lifting objects weighing 10 pounds appears stable across

the two samples. The coefficient for lifting in 1991/1992 is $309 (significant at the 0.01 level) and in 1992/1993 it is $250 (significant at the 0.01 level).

These two samples also show instability in the demographic variables. In the first year, only age less than 65 is significant, while all age categories are significant and have sizable coefficients in the second year. Gender remains significant in both years, but the coefficient increases substantially from $755 in 1991/1992 to $1201 in 1992/1993. These models provide the highest adjusted R-squares of all those tested in this chapter: 0.0280 in the first year, and 0.0340 in the second.

Table A-13 reports findings for the models incorporating the ADL/IADL scale. As might be expected, IADLs only drop out in both years. Of the ADL/IADL ratings, 5-6 ADLs is the only one that remains significant in both years. The physical impairments, walking 2-3 blocks and lifting 10 pounds remain significant in both models as they have in every model except that reported in Table A-12 where walking dropped out.

A.4 Conclusions

Our analyses lead to several conclusions. First, there is considerable redundancy in the disability measures for predicting future Medicare expenditures. We observe similar R-squares for many of the models, and also that combining measures does not greatly increase predictive power. This suggests that policy makers can use criteria other than predictive power to select a set of risk adjusters related to disability. The options are to focus on parsimony, gameability, or concurrent use for other purposes. The parsimony criterion

suggests using those variables that are easiest and least costly to collect. For example, as we have seen in our analyses, the two questions about physical impairments explain almost as much variance as a detailed series of ADL and IADL questions, suggesting that the former might be preferred to the latter. The gameability criteria are objectivity and freedom from manipulation or varying interpretation. For example, the IADL "meal preparation" may be more sensitive to gender roles and less indicative of objective physical health than the questions about walking and lifting. Concurrent uses might include medical treatment outcome measurement, in which case the inclusion of any of these disability measures in an outcome-oriented survey would permit risk adjustment at a low marginal cost.

Our other conclusions relate to variable construction if ADL or ADL and IADL items are included as risk adjusters. Based on these analyses, use of an ADL or ADL/IADL scale (count) seems preferable to use of individual ADLs and IADLs as risk adjusters. Given the small MCBS sample, use of a scale decreases the risk of overfitting the model. There is only a modest decline in R-square from substituting scales for individual functional status measures, and the coefficients of scale models are more stable.

We also recommend the use of ADLs defined by report of difficulty rather than receipt of help. This suggestion is supported by the R-squares, but as importantly, is based on a priori considerations. As reported in Table A-2, only 25 percent of the 1991 sample and 20 percent of the 1992 sample report receiving help with at least one ADL, while 47 percent of the 1991 sample and 40 percent of the 1992 sample report difficulty with at least one ADL. Therefore, the use of difficulty creates adjustments over a broader portion of the Medicare population,

which we believe is appropriate for a payment system intended for the full Medicare population. As we discuss in Chapter 4, there are problems related to the incentive structure and to the endogeneity of help. We believe that it is inappropriate for a payment model to use a measure of impairment that is confounded by availability of help and the provision of care and that the incentives that follow report of difficulty are more desirable. Use of report of difficulty will focus health plans on identifying persons with difficulty and addressing the underlying problems to reduce potential adverse outcomes.

TABLE A-1

FREQUENCY DISTRIBUTION FOR ACTIVITIES OF DAILY LIVING AND INSTRUMENTAL ACTIVITIES OF DAILY LIVING

	1991			1992		
Activities of Daily Living	Percent Reporting Difficulty	Percent Reporting Receipt of Help	Percent of Those with Difficulty Who Receive Help	Percent Reporting Difficulty	Percent Reporting Receipt of Help	Percent of Those with Difficulty Who Receive Help
Bathing	24.0 %	17.2 %	71.7 %	21.2 %	15.9 %	75.1 %
Dressing	16.7	13.4	80.4	15.4	12.5	81.0
Walking	39.7	15.4	38.9	33.5	12.3	36.6
Transfers	23.9	12.1	50.7	20.5	10.5	51.2
Toileting	13.2	8.0	61.0	11.4	7.6	66.4
Eating	8.0	4.9	61.1	6.8	4.8	70.8
Instrumental Activities of Daily Living						
Light Housework	25.4 %	21.4 %	84.0 %	22.5 %	19.9 %	88.4 %
Heavy Housework	47.1	38.0	80.6	43.7	35.0	80.2
Meal Preparation	21.5	18.8	87.3	20.1	18.3	90.7
Shopping	28.8	25.7	89.3	26.3	23.9	90.9
Money Management	18.4	16.7	90.9	18.1	16.7	92.5
Telephone	16.1	8.7	54.3	14.7	8.3	56.4

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991) and 4(1992).

nsk\report\Apndxata xls\TabA-1\ss

TABLE A-2

FREQUENCY DISTRIBUTION OF ADL/IADL SCORE

	1991		1992	
ADL/IADL Score	Percent Reporting Difficulty	Percent Reporting Receipt of Help	Percent Reporting Difficulty	Percent Reporting Receipt of Help
5-6	9.5 %	6.2 %	8.7 %	6.2 %
3-4	10.9	5.8	8.9	4.9
1-2	26.8	12.6	22.6	9.2
IADL only	15.2	23.2	16.0	22.5
None	37.6	52.1	43.8	57.2

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991) and 4(1992).

TABLE A-3

PERCENT REPORTING PHYSICAL IMPAIRMENTS

Physical Impairments	1991 Percent Reporting Difficulty	1992 Percent Reporting Difficulty
Lower Body		
Walking 2-3 blocks	48.1 %	51.6 %
Stooping/Kneeling	70.2	71.6
Lifting	43.5	47.7
Upper Body		
Reaching	32.6	35.6
Writing	32.1	33.8

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991) and 4(1992).

TABLE A-4

DISABILITY MODEL ADJUSTED R-SQUARES

	1991 Characteristics/ 1992 Payments		1992 Characteristics/ 1993 Payments	
	Receives Help[a]	Reports Difficulty[b]	Receives Help[a]	Reports Difficulty[b]
Individual ADLs and IADLs	0.0207	0.0248	0.0262	0.0274
ADLs/IADLs Scale[c]	0.0191	0.0200	0.0242	0.0240
Individual ADLs Only	0.0160	0.0216	0.0218	0.0238
Physical Impairments	-	0.0192	-	0.0291
Individual ADLs and Physical Impairments	0.0254	0.0258	0.0316	0.0301
ADL/IADL Scale and Physical Impairments	0.0255	0.0249	0.0311	0.0296
Degree of Social Function Limitation	-	0.0191	-	0.0274
Any Social Function Limitation	-	0.0143	-	0.0198
ADL/IADL Scale and Social Function Limitation	0.0224	0.0227	0.0293	0.0281
Individual ADLs and Social Function Limitation	-	0.0242	-	0.0287
Individual ADLs and IADLs, Physical Impairments and Social Function Limitation	-	0.0280	-	0.0340
ADL/IADL Scale, Physical Impairments and Social Function Limitation	-	0.0262	-	0.0317

[a] Dichotomous model where 1=reports difficulty and receipt of help

[b] Dichotomous model where 1=reports difficulty including both those who receive help and who don't receive help.

[c] Includes an IADL-only category.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), and 7(1993).

rsk\report\Apndxata\TabA-4\mel

TABLE A-5

REGRESSION OF MEDICARE PAYMENTS ON INDIVIDUAL ADLs AND IADLs

	1991 CHARACTERISTICS/1992 PAYMENTS						1992 CHARACTERISTICS/1993 PAYMENTS					
	Help Cut Off			Difficulty Cut Off			Help Cut Off			Difficulty Cut Off		
	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic
INTERCEPT	2,363	206	11.50 ***	1,928	216	8.92 ***	2,232	207	10.79 ***	1,978	214	9.23 ***
AGE (65-74 OMITTED)												
0-64	-1,107	446	-2.48 ***	-1,426	446	-3.19 ***	681	334	-2.04 **	-860	335	-2.56 **
75-84	484	263	1.84	398	263	1.51	1,134	264	4.29 ***	1,037	265	3.91 ***
85+	560	423	1.32	423	425	0.99	1,164	369	3.15 ***	1,107	372	2.97 ***
MALE	607	240	2.53	722	243	2.96 ***	1,067	226	4.73 ***	1,194	228	5.23 ***
IADLs												
TELEPHONE	-553	552	-1.00	-589	415	-1.42	-560	502	-1.11	-1,060	401	-2.64 ***
LIGHT HOUSEWORK	359	495	0.73	220	447	0.49	284	486	0.58	388	447	0.87
HEAVY HOUSEWORK	1,217	320	3.80 ***	1,117	311	3.59 ***	1,535	311	4.93 ***	1,177	299	3.94 ***
MEAL PREPARATION	577	583	0.99	1,294	522	2.48 **	671	533	1.26	1,263	494	2.55 **
SHOPPING	1,731	434	3.99 ***	884	438	2.02 **	708	409	1.73 *	414	407	1.02
MONEY MANAGEMENT	-967	498	-1.94 *	-470	470	-1.00	-961	437	-2.20 **	-673	420	-1.60
ADLs												
BATHING	1,445	572	2.5[illegible] **	892	437	2.04 **	1,094	569	1.92 *	840	445	1.89 *
DRESSING	81	654	0.12	598	513	1.17	937	622	1.51	802	498	1.61
EATING	468	788	0.59	459	542	0.85	-1,730	709	-2.44 **	-155	550	-0.28
TRANSFERING	-107	628	-0.17	-498	394	-1.26	1,116	600	1.86 *	-153	392	-0.39
WALKING	1,229	459	2.68 ***	1,710	324	5.27 ***	805	474	1.70 *	1,132	319	3.55 ***
TOILETING	-72	769	-0.09	-103	531	-0.19	-207	706	-0.29	287	518	0.56
R SQUARED	0.0221			0.0262			0.0277			0.0289		
ADJUSTED R SQUARED	0.0207			0.0248			0.0262			0.0274		
OBSERVATIONS	10,893			10,893			10,532			10,532		
F	15.33 ***			18.26 ***			18.76 ***			19.56 ***		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor.

*** Statistically significant at 1 percent level.

** Statistically significant at 5 percent level.

* Statistically significant at 10 percent level.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), and 7(1993).

TABLE A-6

SIGNIFICANT VARIABLES IN THE INDIVIDUAL ADLs AND IADLs MODELS

	1991 CHARACTERISTICS/1992 PAYMENTS		1992 CHARACTERISTICS/1993 PAYMENTS	
	Help Cut Off	Difficulty Cut Off	Help Cut Off	Difficulty Cut Off
ADLs	Bathing ** Walking ADL*** Dressing**	Bathing** Walking ADL***	(-) Eating** Transfering* Walking ADL* Bathing*	Walking ADL*** Bathing*
IADLs	Heavy Housework*** Shopping for Personal Items*** (-) Money Management*	Heavy Housework*** Meal Preparation** Shopping for Personal Items**	Heavy Housework*** Shopping for Personal Items* (-) Money Management**	Heavy Housework*** Meal Preparation** (-) Telephone**

*** Statistically significant at 1 percent level.
** Statistically significant at 5 percent level.
* Statistically significant at 10 percent level.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), and 7(1993).

Risk\report\Apndxata xls\TabA-6\nd

TABLE A-7

REGRESSION OF MEDICARE PAYMENTS ON INDIVIDUAL ADLs

	1991 CHARACTERISTICS/1992 PAYMENTS						1992 CHARACTERISTICS/1993 PAYMENTS					
	Help Cut Off			Difficulty Cut Off			Help Cut Off			Difficulty Cut Off		
	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic
Intercept	2,773	194	14.26 ***	2,251	203	11.08 ***	2,598	199	13.05 ***	2,267	204	11.09 ***
Age (65-74 omitted)												
0-64	-456	430	-1.06	-918	432	-2.13 **	-264	320	-0.83	-547	323	-1.70 *
75-84	644	262	2.46 **	484	262	1.84 *	1,302	263	4.96 ***	1,108	264	4.20 ***
85+	945	417	2.27 **	624	417	1.50	1,467	362	4.05 ***	1,206	364	3.32 ***
Male	313	236	1.33	418	237	1.77 *	819	223	3.68 ***	928	223	4.16 ***
ADLs												
Bathing	2,533	530	4.78 ***	1,679	413	4.06 ***	2,172	520	4.17 ***	1,570	418	3.76 ***
Dressing	465	642	0.72	1,097	493	2.23 **	1,074	613	1.75 *	1,057	486	2.18 **
Eating	261	778	0.34	526	531	0.99	-2,036	697	-2.92 ***	-469	535	-0.88
Transferring	237	626	0.38	-259	391	-0.66	1,412	599	2.36 **	85	389	0.22
Walking	2,013	446	4.52 ***	2,294	304	7.54 ***	1,332	466	2.86 ***	1,748	298	5.86 ***
Toileting	-287	765	-0.38	88	527	0.17	-469	702	-0.67	245	513	0.48
R Squared	0.0169			0.0225			0.0227			0.0247		
Adjusted R Squared	0.0160			0.0216			0.0218			0.0238		
Observations	10,893			10,893			10,532			10,532		
F	18.69 ***			25.05 ***			24.45 ***			26.66 ***		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor

*** Statistically significant at 1 percent level.
** Statistically significant at 5 percent level.
* Statistically significant at 10 percent level.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), and 7(1993).

TABLE A-8

REGRESSION OF MEDICARE PAYMENTS ON ADL/IADL SCALE

	1991 CHARACTERISTICS/1992 PAYMENTS						1992 CHARACTERISTICS/1993 PAYMENTS					
	Help Cut Off			Difficulty Cut Off			Help Cut Off			Difficulty Cut Off		
Variable	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic
INTERCEPT	2,256	209	10.81 ***	1,978	223	8.86 ***	2,223	208	10.69 ***	1,953	219	8.92 ***
AGE (65-74 OMITTED)												
0-64	-1,271	443	-2.87 ***	-1,125	440	-2.55 **	-961	336	-2.86 ***	-872	334	-2.61 ***
75-84	500	263	1.90 *	447	264	1.70 *	1,082	265	4.09 ***	1,003	266	3.77 ***
85+	662	418	1.58	721	416	1.73 *	1,086	366	2.97 ***	1,123	365	3.08 ***
MALE	530	238	2.22 **	477	237	2.01 **	991	224	4.42 ***	1,006	224	4.48 ***
FUNCTIONAL STATUS (None omitted)												
5-6 ADLs	5,680	588	9.65 ***	5,589	488	11.45 ***	4,511	513	8.79 ***	4,882	446	10.94 ***
3-4 ADLs	4,564	575	7.94 ***	3,517	428	8.23 ***	5,150	536	9.61 ***	3,648	418	8.74 ***
1-2 ADLs	2,985	394	7.57 ***	2,537	294	8.63 ***	2,960	399	7.42 ***	2,336	290	8.06 ***
IADLs Only	2,125	303	7.02 ***	1,129	357	3.17 ***	1,774	282	6.30 ***	1,239	324	3.83 ***
R-Squared	0.0198			0.0207			0.0249			0.0247		
Adjusted R Squared	0.0191			0.0200			0.0242			0.0240		
Observations	10,893			10,893			10,532			10,532		
F	27.49 ***			28.71 ***			33.53 ***			33.33 ***		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor

*** Statistically significant at 1 percent level.
** Statistically significant at 5 percent level
* Statistically significant at 10 percent level.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), and 7(1993).

TABLE A-9

REGRESSION OF MEDICARE PAYMENTS ON PHYSICAL IMPAIRMENTS

	1991 CHARACTERISTICS/ 1992 PAYMENTS			1992 CHARACTERISTICS/ 1993 PAYMENTS		
	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic
Intercept	1,642	262	6.28 ***	325	286	1.14
Age (65-74 omitted)						
0-64	-949	435	-2.18 **	-761	323	-2.35 **
75-84	449	264	1.70 *	926	264	3.50 ***
85+	997	412	2.42 **	959	362	2.65 ***
Male	681	241	2.83 ***	1,141	225	5.07 ***
Physical Impairments						
Stooping	-170	288	-0.59	105	108	0.97
Walking 2-3 blocks	1,892	286	6.61 ***	658	97	6.79 ***
Lifting	1,678	295	5.69 ***	382	102	3.75 ***
Reaching	344	288	1.20	75	109	0.69
Writing	245	278	0.88	129	122	1.06
R-Squared	0.0200			0.0299		
Adjusted R-Squared	0.0192			0.0291		
Observations	10,860			10,531		
F	24.55 ***			36.03 ***		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor.

*** Statistically significant at 1 percent level.

** Statistically significant at 5 percent level.

* Statistically significant at 10 percent level.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), and 7(1993).

TABLE A-10

REGRESSION OF MEDICARE PAYMENTS ON DEGREE OF SOCIAL FUNCTION LIMITATION

	1991 CHARACTERISTICS/ 1992 PAYMENTS			1992 CHARACTERISTICS/ 1993 PAYMENTS		
Variable	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic
Intercept	2,332	207	11.26 ***	2,196	209	10.53 ***
Age (65-74 omitted)						
0-64	-841	434	-1.94 *	-757	326	-2.32 **
75-84	641	262	2.45 **	1,189	263	4.53 ***
85+	1,486	404	3.68 ***	1,925	347	5.55 ***
Male	361	236	1.53	845	222	3.80 ***
Social Function Limited[1] (None of the time omitted)						
All of the time	5,343	452	11.82 ***	5,508	437	12.61 ***
Most of the time	2,875	390	7.36 ***	3,323	366	9.08 ***
Some of the time	1,709	301	5.69 ***	1,386	271	5.12 ***
R Squared	0.0197			0.0280		
Adjusted R Squared	0.0191			0.0274		
Observations	10,876			10,519		
F	31.24 ***			43.22 ***		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor.

*** Statistically significant at 1 percent level.

** Statistically significant at 5 percent level.

* Statistically significant at 10 percent level.

[1]Based on MCBS question "How much of the time during the last month has (your/SP's) health limited (your/SP's) social activities, like visiting with friends or close relatives?"

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), and 7(1993).

TABLE A-11

REGRESSION OF MEDICARE PAYMENTS ON ANY SOCIAL FUNCTION LIMITATION

	1991 CHARACTERISTICS/ 1992 PAYMENTS			1992 CHARACTERISTICS/ 1993 PAYMENTS		
Variable	Coefficient	Standard Error	t Statistic	Coefficient	Standard Error	t Statistic
Intercept	2,291	207	11.05 ***	2,119	209	10.13 ***
Age (65-74 omitted)						
0-64	-723	434	-1.67 *	-558	326	-1.71 *
75-84	723	262	2.76 ***	1,300	263	4.94 ***
85+	1,717	403	4.26 ***	2,215	346	6.40 ***
Male	366	237	1.55	819	223	3.67 ***
Social Function[1]	2,714	245	11.06 ***	2,581	229	11.29 ***
R Squared	0.0148			0.0203		
Adjusted R Squared	0.0143			0.0198		
Observations	10,893			10,532		
F	32.74 ***			43.52 ***		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor.

*** Statistically significant at 1 percent level.

** Statistically significant at 5 percent level.

* Statistically significant at 10 percent level.

[1]Range of responses collapsed. 1=social activities limited by health in the last month any amount of time (some of the time to all of the time). 0=none of the time.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), and 7(1993).

TABLE A-12

BACKWARDS STEPWISE REGRESSION OF MEDICARE PAYMENTS ON INDIVIDUAL ADLs AND IADLs, PHYSICAL LIMITATIONS AND SOCIAL FUNCTION LIMITATIONS - FINAL MODELS

	1991 CHARACTERISTICS/ 1992 PAYMENTS			1992 CHARACTERISTICS/ 1993 PAYMENTS		
	Reports Difficulty			Reports Difficulty		
	Coefficient	Standard Error	F	Coefficient	Standard Error	F
Intercept	959	275	12.12 ***	884	257	11.86 ***
Age (65-74 omitted)						
0-64	-1,638	441	13.76 ***	-1,177	330	12.71 ***
75-84	307	262	1.37	852	264	10.37 ***
85+	174	420	0.17	772	370	4.34 **
Male	755	242	9.73 ***	1,201	225	28.51 ***
ADLs						
Bathing	805	409	3.87 **	965	384	6.31 **
Walking (ADL)	347	114	9.19 ***	502	92	29.71 ***
Transferring	-627	373	2.83 *			
IADLs						
Meal Preparation	1,302	419	9.67 ***	1,379	397	12.10 ***
Heavy Housework	552	318	3.02 *			
Telephone				-1,124	370	9.23 ***
Physical Impairments						
Walking (2-3 blocks)	864	374	5.35 **			
Lifting	309	106	8.54 ***	250	98	6.53 **
Social Function	922	288	10.28 ***	1,277	255	25.16 ***
R Squared	0.0291			0.0348		
Adjusted R Squared	0.0280			0.0340		
Observations	10,893			10,532		
F	27.15 ***			37.97 ***		
Dropped Variables	Light Housework Writing Toileting Reaching Stooping Eating Money Management Shopping Dressing Telephone			Stooping Shopping Light Housework Reaching Walking (2-3 blocks) Eating Writing Toileting Transferring Money Management Dressing Heavy Housework		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor.

*** Statistically significant at 1 percent level.

** Statistically significant at 5 percent level.

* Statistically significant at 10 percent level.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), and 7(1993).

TABLE A-13

BACKWARD STEPWISE REGRESSION OF MEDICARE PAYMENTS ON ADL/IADL SCALE, PHYSICAL LIMITATIONS AND SOCIAL FUNCTION LIMITATIONS - FINAL MODELS

	1991 CHARACTERISTICS/ 1992 PAYMENTS			1992 CHARACTERISTICS/ 1993 PAYMENTS		
	Reports Difficulty			Reports Difficulty		
	Coefficient	Standard Error	F	Coefficient	Standard Error	F
Intercept	731	265	7.58 ***	733	255	8.25 ***
Age (65-74 omitted)						
0-64	-1,334	435	9.41 ***	-1,007	326	9.52 ***
75-84	369	262	1.98	881	264	11.12 ***
85+	473	415	1.30	887	363	5.97 **
Male	664	238	7.76 ***	1,129	224	25.31 ***
Functional Status						
5-6 ADLs	1,751	532	10.83 ***	1,119	458	5.97 **
1-2 ADLs	513	293	3.06 *			
Physical Impairments						
Walking (2-3 blocks)	519	99	27.45 ***	571	91	39.52 ***
Lifting	418	101	17.03 ***	325	96	11.38 ***
Social Function	1,115	282	15.62 ***	1,311	255	26.40 ***
R Squared	0.0270			0.0324		
Adjusted R Squared	0.0262			0.0317		
Observations	10,893			10,532		
F	33.52 ***			44.08 ***		
Dropped Variables	Writing Reaching Stooping 3-4 ADLs IADLs only			Writing Reaching Stooping 3-4 ADLs 1-2 ADLs IADLs only		

NOTE: Dependent variable annualized and the MCBS sample weights were multiplied by the inverse of the annualization factor.

*** Statistically significant at 1 percent level.

** Statistically significant at 5 percent level.

* Statistically significant at 10 percent level.

SOURCE: Medicare Current Beneficiary Survey, Rounds 1(1991), 4(1992), and 7(1993).

Appendix B

SF-36-Like Model

B

Construction of Simulated SF-36 Health Status Scales Using the Medicare Current Beneficiary Survey

In this appendix, we explain how we simulated SF-36 health status scales using the MCBS. We identified MCBS proxies for individual SF-36 variables, and used the MCBS/SF-36 variables to create scales based on SF-36 scoring methodologies (Ware et al., 1993).

The SF-36 uses a scaling approach to combine the effects of specific variables into domains of health status measurement. These scales are Physical Functioning, Role-Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role-Emotional, Mental Health, and Reported Health Transition. The SF-12 uses the 12 out of the 36 items that account for the most variance in the two core concepts, physical health and mental health.

The MCBS has items from several of these scales. None are exactly equivalent. Some differ in the wording of the question, others in the scaling. We were able to simulate the following four SF-36 scales:

- General health;
- Social functioning;
- Physical functioning; and
- Role-physical.

Table B-1 lists the items in the four SF-36 scales that we simulated, the items included in the SF-12, a shortened 12 item variant of the SF-36, and the MCBS item we used as an

equivalent. For example, the MCBS asks about self-rated health "Compared to others your age...". The SF-36 does not qualify the question by age. The social function question in the SF-36 asks due to physical or emotional problems, the MCBS only stipulates due to health problems. Within the physical function questions, the MCBS offers more possible responses to each item than does the SF-36.

Although worded differently, the MCBS does have the key questions from the SF-36 used for General Health and Social Function, as indicated by the use of these questions in the SF-12 as the sole indicators of the SF-36 scales. For the physical functioning scale, the MCBS has five of the ten SF-36 variables, six if we use the Walk-ADL as a substitute for walking one block. According to the SF-36 manual, missing values for individual items in the physical functioning scale are imputed at the level of the average response to the completed items as long as at least half the items in the scale are present. Hence, if we use the physical functioning variables we do have in the MCBS to approximate the SF-36 responses, we can generate Physical Functioning scores.

Higher scores in the SF-36 are associated with better health. For some items this requires recoding. Some items are scored in a simple linear fashion, some are weighted, and some are transformed to a standardized scale. Because higher scores are associated with better health, predictive models based on SF-36 scoring will have a higher intercept and the coefficients will be negative. In other words, the coefficients will represent reductions in costs related to SF-36 scores.

In this appendix we describe the MCBS proxies we chose for individual SF-36 variables based on a crosswalk between the two. We converted the MCBS response choices to approximate SF-36 response choices where the two differ.

1.0 General Health

Both the SF-36 and MCBS offer a five point response scale from poor to excellent. However, the stems of the question differ. The MCBS asks "Compared to others your age..." while the SF-36 does not place this modifier on the question.

The self-rated health question in the SF-36 is not scored in a simple linear fashion, because the differences among the responses ranging from good to excellent were found to be smaller than those ranging from good to poor. The recommended scoring, which we used with the MCBS variable, is:

Excellent	5.0
Very Good	4.4
Good	3.4
Fair	2.0
Poor	1.0

2.0 Social Function

The wording and rating scales differ in the two questionnaires. The MCBS also varies between rounds. The SF-36 asks "During the past 4 weeks, how much of the time has your

physical health or emotional problems interfered with your social activities (like visiting with friends, relatives, etc)?" and offers a five point response scale. The MCBS equivalent is less specific about physical health or emotional problems, asking "How much of the time during the past month has (your/sp's) **health limited** (your/sp's) social activities, like visiting with friends or close relatives?" In Round 1, institutional residents are offered a six point response scale and community residents a 4 point scale. In the later rounds, all respondents are offered a four point response scale. To proxy for the SF-36 responses, we first collapsed the Round 1 institutional responses to a 4 point scale equivalent to that used in all subsequent rounds. Second, we collapsed the MCBS responses "some of the time" and "a little of the time" to be equivalent to the SF-36 response "some of the time".

This scale is scored in a simple linear fashion. In the case of conversion to the MCBS, we collapsed the responses "a little of the time" and "some of the time" to develop a four point scale as in the SF-36. The SF-36 scoring of the responses is a simple count as indicated below.

Social Activities limited	All of the time	1
	Most of the time	2
	Some of the time	3
	None of the time	4

3.0 Physical Functioning

The SF-36 has 3 possible responses for each of the 10 items in this scale (see table), referring to the degree of limitation, scored as indicated:

No, not limited at all 3

Yes, limited a little 2

Yes, limited a lot 1

The responses to these items are added up to create a raw score. This score is then transformed as follows:

$$\text{Transformed Scale} = \frac{\text{(Actual raw score - lowest possible raw score)}}{\text{Possible raw score range}} \times 100$$

In this scale, 10 is the lowest possible raw score, and the possible raw score range is 20.

We were able to find MCBS equivalents to 5 of the 10 SF-36 physical functioning items, as indicated in the table. According to the SF-36 scoring manual (Ware et al., 1993), the physical functioning scale can be created with missing data on as many as 5 of the 10 items, with an adjustment for the missing items. The formula we used to compute our simulated physical functioning scale was:

$$\frac{\text{(heavy housework score+walking 2-3 blocks score+ lifting score+stoop score+bathe or dress score + average of the preceding x 5) - 10}}{20} \times 100$$

We now describe how we scored each of the items in this scale.

3.1 Simulating SF-36 responses for the physical functioning variables

3.1.1 Walking, lifting/carrying, stooping/kneeling

Difficulty with lifting or carrying 10 pounds, stooping/kneeling, and difficulty walking 2-3 blocks are in both questionnaires, although the wording differs slightly. The MCBS offers

a five point response scale, the SF-36 has a 3 point response scale. We converted MCBS responses to three-level SF-36 equivalents as follows for the "walking", "lifting or carrying", and "stooping/kneeling" variables:

Score = 3, not limited at all
if MCBS responses include: No difficulty

Score = 2, yes, limited a little
if MCBS response is: a little or some difficulty performing

Score =1, yes, limited a lot
if MCBS response is: a lot of difficulty or unable to perform

3.1.2 ADLs

The SF-36 asks "Does your health now limit you in bathing or dressing yourself, if so how much?" As indicated in Section 3.1.1, the response options are: limited a lot, limited a little, or not limited. The MCBS has separate bathing and dressing questions, and a more complicated set of responses (as described in Section 3.1 above). We collapsed the responses for these two ADLs in the MCBS to create one score capturing both bathing and dressing as follows:

Score = 3, not limited at all
if MCBS responses include: No difficulty with either bathing or dressing, or don't bathe or dress not due to a health reason.

Score = 2, yes, limited a little
if MCBS response is: difficulty performing, but do not receive help, with either bathing or dressing.

Score =1, yes, limited a lot
if MCBS response is: unable to either bath or dress due to a health reason, or receive help with either bathing or dressing.

3.1.3 Difficulty with Heavy Housework

The SF-36 asks if the respondent has difficulty with "moderate activities", such as moving a table, pushing a vacuum cleaner, bowling or playing golf. We consider the MCBS question about difficulty with "heavy housework" to be a reasonable equivalent. We collapsed the MCBS responses to a three point scale similar to the SF-36 ratings as follows:

Score = 3, not limited at all
if MCBS responses include: No difficulty, or Doesn't do for nonhealth reason

Score = 2, yes, limited a little
if MCBS response is: Reports Difficulty but does not receive help

Score =1, yes, limited a lot
if MCBS response is: Receives help, or Doesn't do for a health reason

Those who responded that they don't do heavy housework for a health reason are coded as "limited a lot", the same response as for those who reported difficulty and receive help. Categorizing those who reported that they don't do heavy housework, but did not attribute the lack of this activity to a health reason, is more difficult. After extensive analyses to identify similarities and differences between those who "Don't do heavy housework" without citing a health reason and other groups, we chose to recode this response as having no difficulty. This group was found to be significantly less ADL impaired than those who reported "doesn't do due to health," less likely to be among the younger disabled or 85 and

over group, more likely to be male, and less likely to report fair or poor general health. In comparison to the "no difficulty" group, while slightly older, slightly more impaired in ADLs, social function, general health and having higher rates of chronic disease, their Medicare expenditures were not significantly different. Institutional residents are not asked this question in the MCBS. We imputed a value of "limited a little" to all institutionalized respondents.

4.0 Role Physical

The SF-36 includes four questions that contribute to a subscale called "Role-Physical." The purpose of this scale is to identify the impact of health status on work or other usual activities. These questions ask if, due to physical health, the respondent has "cut down on the amount of time spent on work or other regular daily activities; accomplished less than you would like; were limited in the kind or amount of work or other regular daily activities; had difficulty performing work or other regular daily activities" in the past four weeks. For the vast majority of elderly and disabled Medicare beneficiaries, paid "work" is not a regular daily activity. We hypothesized that the instrumental activities of daily living (IADL) questions on the MCBS provide an appropriate measurement of "regular daily activities" for the Medicare population. Hence, a scale analogous to the SF-36 "role physical" scale can be constructed from the MCBS IADL responses. Our equivalence hypothesis has not been empirically verified or validated.

In the SF-36 role-physical scale, each of the four items are scored either 1 (yes, has cut down on work or activities, etc) or 2 (no, has not cut down, accomplished less, been limited, or had difficulty). The scores of these items are summed, so that the range of possible values for this scale is from 4 (least impaired) to 8 (most impaired).

In our simulated role physical scale, we followed the same approach to scoring, substituting IADL measures for the SF-36 items (i.e., an IADL response of "difficulty" = 1, "no difficulty" = 2). We did not use difficulty with heavy housework since it was already part of our simulated physical function scale. Based on available variables, we chose to use difficulty with light housework, shopping, money management and using the telephone.[1] While these are not equivalent to the series in the SF-36 that measures different types of difficulty one might experience with performing regular activities, the scale we created does measure extent of impairment in typical activities that might be routinely performed by the elderly and disabled.[2] Thus, we hypothesize that it is conceptually analogous to the SF-36 role physical scale, although differing in details of measurement.

[1] Facility residents were not asked about "light housework". We imputed difficulty to all facility residents for this question. If the response to the IADL was "don't do", "difficulty" was coded if it was "because of a health reason", "no difficulty" was coded if it was "not because of a health reason". We did not employ the IADL for "preparing meals" in our role physical scale, because many men reported that they don't do it not because of a health reason. Also, it would have to be imputed for facility respondents.

[2] A different between the SF-36 and the MCBS IADLs is that the SF-36 refers to "the past four weeks", while the MCBS question does not specify a time frame.

Table B-1

Crosswalk Between MCBS and SF-36 and SF-12 Items

SF-36 Scale	SF-12 Items	MCBS Items
General Health		
In general, is your health: excellent, very good, good, fair, poor	In general, is your health: excellent, very good, good, fair, poor	In general, compared to others your age, ...
I seem to get sick a little earlier than other people		
I am as health as anybody I know		
I expect my health to get worse		
My health is excellent		
Social Functioning		
During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities like visiting with friends or relatives?	During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities like visiting with friends or relatives?	How much of the time during the past month has your health limited your social activities, like visiting with friends or close relatives?
During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors or groups?		
Physical Functioning		
Vigorous activities		
Moderate activities	Moderate activities	Heavy Housework
Carry groceries		Lift 10 pounds
Climb several flights	Climb several flights	
Climb one flight		
Bend, kneel, stoop		Bend, kneel, stoop
Walk a mile		
Walk several blocks		Walk 2-3 blocks
Walk one block		
Bathe or dress		Bathe, dress

Appendix C

Sample Sizes Necessary for Payment Accuracy

C

Sample Sizes Necessary for Payment Accuracy

When HCFA sets capitated payments to HMOs for treating Medicare beneficiaries, there are two issues of accuracy:

1) setting an HMO's payment amount accurately given a certain payment formula; and

2) predicting the actual mean expenditures of an HMO's enrollees accurately using the payment formula.

With survey-based risk adjusters, the first issue is an issue of sampling, that is, from how many of an HMO's enrollees do health status measures have to be collected to accurately determine payment for the HMO? The necessary sample size depends on the variance of the expenditures predicted by the payment formula, such as the AAPCC, or a revised formula incorporating health status measures. The second issue is a function of the variance of <u>actual</u> expenditures, because we are comparing predicted to actual expenditures.

Because the variance of predicted expenditures is much smaller than the variance of actual expenditures, accuracy in the first sense requires a much smaller sample size than accuracy in the second sense. Table 1 shows representative sample sizes necessary to set payments accurately for an HMO under alternative models. A certain number of an HMO's enrollees would be sampled through a survey sampling methodology to determine the average (mean) characteristics of an HMO's enrollees. Payment would then be based on these average characteristics. It is unnecessary to determine the characteristics of each enrollee

because it is sufficient to make a single total payment to the HMO. As can be seen from the table, the required sample sizes are quite modest. Required sample sizes for payment models incorporating health status are larger than required for the AAPCC[1] because they have a greater variance of predicted spending. They are larger the greater the required precision in HMO payment, e.g., for a plus or minus 2% error with 99% confidence as opposed to a plus or minus 10% error with 90% confidence.

Table 2 shows sample sizes necessary to achieve accuracy in the second sense, that is, a given deviation of mean expenditures predicted by the payment formula from actual mean expenditures. Because actual medical expenditures are highly variable, the required sample sizes here are much larger. However, the required sample sizes are smaller for the more highly predictive models incorporating health status. This is because the models incorporating health status predict expenditures better, so few observations are necessary to achieve a given level of precision. Obviously, the sample sizes necessary to achieve a close fit between predicted and actual expenditures may be larger than the enrollment size of smaller HMOs. This means that even if all their enrollees were surveyed, one could not expect a very close relationship between predicted and actual expenditures. In other words, smaller HMOs will be incurring financial risk by participating in Medicare which may need to be controlled through other means (e.g., reinsurance).

We include a technical note to illustrate how the numbers were derived.

[1] Of course, the AAPCC factors (other than institutionalization) can be determined from Medicare administrative records, so surveys are unnecessary. Sample sizes for the demographic AAPCC model are presented for comparative purposes.

TECHNICAL NOTE

Sample Size Requirements

Both Table C-1 and Table C-2 use the same theoretical formula. Let predicted spending for the sample of beneficiaries in a hypothetical HMO plan be labeled $\hat{y}$. The following assumptions hold.

(a) $\hat{y}$ is distributed with mean μ and variance σ^2.

(b) The sample mean of $\hat{y}$ (from a sample of size n), or $\bar{\hat{y}}$ is distributed with mean μ and variance σ^2/n.

(c) Using central limit theorem, $\sqrt{n}\cdot\left[\frac{\bar{\hat{y}}-\mu}{\sigma}\right]$ is asymptotically distributed as a normal distribution with mean 0 and variance 1. Thus,

$$z = \sqrt{n}\cdot\left(\frac{\bar{\hat{y}}-\mu}{\sigma}\right)$$

$$\Rightarrow z^2 = n\cdot\left[\frac{\left(\bar{\hat{y}}-\mu\right)^2}{\sigma^2}\right]$$

$$\Rightarrow z^2\sigma^2 = n\cdot\left[\frac{\left(\bar{\hat{y}}-\mu\right)^2}{\bar{\hat{y}}^2}\right]\cdot\bar{\hat{y}}^2$$

Using π to denote the percentage error between the sample mean ($\bar{\hat{y}}$) and the population mean μ:

$$\Rightarrow z^2\,\sigma^2 = n\,\pi^2\,\bar{\hat{y}}^2.$$

$$\Rightarrow n = z^2\,\sigma^2/\pi^2\,\bar{\hat{y}}^2.$$

To estimate σ^2, we used the sample variance of predicted spending in Table C-1 and the Mean Squared Error (MSE) from the various regression models used to estimate the contributions of risk adjustment factors in Table C-2.

TABLE C-1

SURVEY SAMPLE SIZE REQUIRED TO SET HMO PAYMENT WITH SPECIFIED ACCURACY FOR ALTERNATIVE PAYMENT MODELS

	PERCENTAGE DIFFERENCE BETWEEN ACTUAL AND EXACT MEAN PAYMENT			
Payment Model	2%	5%	10%	20%
95% confidence				
Age-gender	362	58	14	4
Self-Rated Health Status	1,788	286	72	18
Functional status	1,913	306	77	19
Self-Reported Chronic Conditions	2,450	392	98	25
Comprehensive Survey	3,324	532	133	33
99% confidence				
Age-gender	626	100	25	6
Self-Rated Health Status	3,085	494	123	31
Functional status	3,302	528	132	33
Self-Reported Chronic Conditions	4,229	677	169	42
Comprehensive Survey	5,737	918	229	57
90% confidence				
Age-gender	255	41	10	3
Self-Rated Health Status	1,260	201	50	13
Functional status	1,348	216	54	13
Self-Reported Chronic Conditions	1,726	276	69	17
Comprehensive Survey	2,342	375	94	23

NOTE: Table entries are numbers of Medicare beneficiaries enrolled in an HMO. "Exact" payment assumes an infinite population of HMO enrollees whose characteristics are all known and used to compute payments. "Actual" payment is that computed from the specified sample of HMO enrollees.

SOURCE: HER computations from the 1991/1992 Medicare Current Beneficiary Survey.

TABLE C-2

SAMPLE SIZE REQUIRED TO PREDICT MEAN ACTUAL EXPENDITURES FOR AN HMO's ENROLLEES WITH THE SPECIFIED ACCURACY

	PERCENTAGE DIFFERENCE BETWEEN ACTUAL AND PREDICTED MEAN EXPENDITURES			
Payment Model	2%	5%	10%	20%
95% confidence				
Age-gender	96,377	15,420	3,855	964
Self-Rated Health Status	94,986	15,198	3,799	950
Functional status	94,861	15,178	3,794	949
Self-Reported Chronic Conditions	94,384	15,101	3,775	944
Comprehensive Survey	93,701	14,992	3,748	937
99% confidence				
Age-gender	166,347	26,616	6,654	1,663
Self-Rated Health Status	163,947	26,232	6,558	1,639
Functional status	163,730	26,197	6,549	1,637
Self-Reported Chronic Conditions	162,907	26,065	6,516	1,629
Comprehensive Survey	161,728	25,876	6,469	1,617
90% confidence				
Age-gender	67,909	10,862	2,716	679
Self-Rated Health Status	66,929	10,705	2,676	669
Functional status	66,840	10,691	2,673	668
Self-Reported Chronic Conditions	66,504	10,637	2,659	665
Comprehensive Survey	66,023	10,560	2,640	660

NOTE: Table entries are numbers of persons.

SOURCE: HER computations from the 1991/1992 Medicare Current Beneficiary Survey.